Journal of Experimental Psychology:
Learning, Memory, and Cognition
1988, Vol. 14, No. 3, 484-494

0278-7393/88/$00.75

The Generation Effect: Support for a Two-Factor Theory

Elliot Hirshman and Robert A. Bjork
University of California, Los Angeles

When a response word bearing an orthographic, acoustic, or semantic relation to a stimulus word is generated rather than read, later recall is enhanced. Such "generation effects" have been attributed to the activation or strengthening of response-specific features in memory and to the activation or strengthening of the relation between a stimulus and response. This series of experiments yields evidence suggesting that both mechanisms are involved. The pattern of interactions in the size of the generation effect across type of recall test (cued or free) cannot be accommodated by any one-factor theory. The results of these experiments also suggest that within-subjects manipulations of read and generate study conditions inflate the apparent size of the effect of generation on a given pair by confounding such pair-specific effects with certain whole-list effects, such as differential attention and output interference.

Research in the last 10 years has provided evidence that items subjects generate are better remembered than items they read. This phenomenon, dubbed the *generation effect* by Slamecka and Graf (1978), has proved to be remarkably robust. It has been found in cued recall (Donaldson & Bass, 1980; Ghatala, 1983; Graf, 1980; Jacoby, 1978; Johnson, Raye, Foley, & Foley, 1981), recognition memory (Gardiner & Rowley, 1984; Glisky & Rabinowitz, 1985; Graf, 1982; Jacoby, 1983; McElroy & Slamecka, 1982; McFarland, Frey, & Rhodes, 1980; Nairne, Pusen, & Widner, 1985; Nairne & Widner, 1987; Payne, Neely, & Burns, 1986; Slamecka & Fevreiski, 1983; Slamecka & Graf, 1978), and free recall (Gardiner & Arthurs, 1982; Gardiner & Hampton, 1985; Gardiner & Rowley, 1984; Johnson et al., 1981; McFarland et al., 1980; Slamecka & Graf, 1978). The effect has been found with a variety of materials (single words, word pairs, sentences), generation rules (associate, rhyme, synonym–antonym, letter and numeral transposition), retention intervals (up to 10 days), and with intentional or incidental learning procedures.

There have been several notable failures to find the effect. Graf (1980) used a sentence generation task and found no generation effect in cued recall when the to-be-generated sentences were anomalous. Donaldson and Bass (1980) used the traditional word-pair method invented by Slamecka and Graf (1978), in which a response word is generated from a stimulus word and some letter cues of the response word, and found no generation effect in cued recall when both read and generate subjects performed a semantic adequacy check. In the semantic adequacy check, they asked subjects to rate the closeness of the semantic relation between the words.

The remaining notable failure to find a generation effect occurs when the response terms are nonsense words or do not correspond to some familiar concept in semantic memory. McElroy and Slamecka (1982), Nairne et al. (1985), Gardiner and Rowley (1984), Gardiner and Hampton (1985), and Payne et al. (1986) have reported this finding. These findings have generally been interpreted to indicate that the activation of specific, preexisting item features, such as lexical attributes, is necessary for the effect to occur.

Single-Factor Theories of the Generation Effect

Many of the theoretical explanations put forward to account for the generation effect have focused on differences in effort or depth of processing (Graf, 1980; McElroy & Slamecka, 1982; McFarland et al., 1980; Slamecka & Graf, 1978). All such explanations, however, appear to be discredited by Slamecka and Graf's (1978) finding that there is no generation advantage (when a rhyme rule is used) in recognition memory for the stimulus term. This finding discredits effort-type theories because one would expect the stimulus term, especially with an associative rule, to be as deeply or as effortfully processed as the response term. Damaging the effort explanations further are the failures to find the effect with nonsense syllables, which are presumably processed effortfully. (See, however, Nairne & Widner, 1987, for a dissenting view on the failure to find the generation effect with nonsense syllables.)

The remaining theories of the generation effect are of two types. The first type focuses on the relations among items. More specifically, such theories claim that the connection from the stimulus term to the response term is strengthened by the act of generation. The second type of explanation focuses on item-specific properties of the response term; the act of generation is assumed to enhance the activation of an item's features. Hereinafter, these two theories are referred to as *relational* and *item-specific* theories.

The foregoing distinction between relational and item-specific information follows that of Hunt and Einstein (1981), except that they used *relational* to denote the abstraction of common features among items, whereas the term is used here to denote the more general relation between the stimulus and the response term. (See Anderson, 1972, for a similar distinction.)

The research reported here was supported in part by Grant 3186 from the Committee on Research, University of California, Los Angeles, to Robert A. Bjork.

Correspondence concerning this article should be addressed to Robert A. Bjork, Department of Psychology, Franz Hall, University of California, Los Angeles, California 90024-1563 or to Elliot Hirshman, Department of Psychology, New York University, New York, New York 10003.

The relational theory emerges from the work of Donaldson and Bass (1980) and Graf (1980). On the basis of data from their experiment requiring subjects to carry out a semantic adequacy check, Donaldson and Bass claimed the generation effect occurs because, following generation, subjects routinely perform such a semantic adequacy check, whereas simple reading does not incorporate such a check. Examining semantic appropriateness presumably strengthens the relation from the stimulus to the response term. McElroy and Slamecka (1982) have criticized this conclusion because the adequacy check raised the scores of both the read and generate subjects, which raises the question of why the generate group received a benefit from the postgeneration adequacy check if they had carried out this check as an intrinsic part of the generation process. The failure to find a generation effect on the stimulus term further damages Donaldson and Bass's hypothesis because the adequacy check hypothesis claims that additional effort or attention is also devoted to the stimulus term.

Graf (1980), on the basis of the failure to find a generation effect in cued recall for anomalous sentences, also claimed that generation strengthens the relations between items. When there are no such relations available, as in the anomalous sentences, the effect disappears. McElroy and Slamecka (1982) also criticized this conclusion. They pointed out that in Graf's sentence-generation paradigm, nothing is actually generated from semantic memory. Such an absence of contact with semantic memory, they claimed, raises questions about whether we can generalize from Graf's sentence-generation paradigm to the traditional paradigm of Slamecka and Graf (1978).

The second type of theory—that item-specific features of the response term mediate the generation effect—was triggered by McElroy and Slamecka's (1982) finding that there is no generation effect when the response term is a nonsense word. According to this *lexical activation hypothesis*, the generation effect occurs because generation produces greater activation of an item's semantic attributes than does reading. The work of Gardiner and his colleagues has indicated that the hypothesis extends to any familiar concept, not just lexical entries. A variant of this hypothesis is Nairne et al.'s (1985) suggestion that the effect is due to the activation of an item's associates: an idea based on their failure to find a generation effect in recognition with low-frequency words.

There is considerable controversy in the literature over the relative roles played by item-specific and relational factors in producing the generation effect. On the basis of recent evidence it now appears that a relational explanation is not sufficient. Glisky and Rabinowitz (1985) found a generation effect in recognition when single words were the to-be-generated items, a result that is clearly a problem for theories based solely on relational factors. Similarly, Gardiner and Hampton (1985) have shown a generation effect in free recall using a letter or number transposition rule in which it seems unlikely that the stimulus term could act as a cue for the response term. Evidence such as this led Gardiner and Hampton to conclude "what is now known about the generation effect strongly suggests that, in essence, the effect is due to item-specific rather than relational processing" (p. 740).

Goals and Methodological Characteristics of the Present Research

Given that a relational explanation is not sufficient to explain all the effects of generation, the question remains as to whether a one-factor item-specific explanation, such as the lexical-activation hypothesis, is itself sufficient. The first goal of our experiments is to answer that question. The traditional Slamecka and Graf (1978) paradigm was implemented in a form designed to maximize the possibility of finding the effects of relational factors, if they exist. Toward that end, related rather than unrelated pairs were used, and the nature of the relation was associative, because such encoding would seem to depend more heavily on relational information than would a rhyme or transposition rule. In addition, both free- and cued-recall tasks were used because it was thought they would be differentially affected by relational and item-specific factors.

If it is necessary to assume that generated items have relational, as well as item-specific, advantages over read items, then the determination of the relative roles played by these two factors on cued-recall and free-recall tests is theoretically important. Consistent with the classic view (Underwood & Schulz, 1960) that cued-recall tests involve response learning and associative learning, we assume that both item-specific and relational factors contribute to the generation effect in cued recall. In free recall, however, we assume that only the item-specific factor causes the generation effect. The bases for this assumption are the findings of Hayes-Roth (1977) and Yekovich and Manelis (1980), which demonstrate that as strings of items become more strongly related, free recall of items within these strings is reduced. Generalizing these results to the paired-associate learning paradigm, we assume that the stronger relation between the stimulus and response which results from generating the response, does not help, and can even hurt, later free recall of the generated response. Consistent with this assumption, responses that are third associates of the stimulus terms are better free recalled than responses that are third associates of the stimulus terms in the present experiments.

Because the second goal of the present research is to test the particular two-factor interpretation of the effects of generation presented earlier, the generation effects of interest are those pair-specific effects attributable to generating a response from a related stimulus, as opposed to reading that stimulus–response pair. Other whole-list effects that arise as a consequence of mixing read and generate items in the same list—that is, within-subjects—we consider to be second-order effects that are not a consequence of generation per se. More specifically, when read and generate items are mixed in a list, there is good reason (see Begg & Snider, 1987, and the present Experiment 4) to believe that there is differential allocation of attention and rehearsal to the generate items, and that such whole-list differential processing of generate items may also subject read items to differential output interference at recall. For that reason, with the exception of Experiment 4, we chose to manipulate read–generate on a between-lists (and between-subjects) basis, rather than follow the far more frequent practice of manipulating that factor on a within-subjects basis.

To put the argument in Poulton's (1982) terms, the generate items are "influential companions" in a within-list design because they alter (impair) the processing of the read items in the list. Poulton was critical of the preference among cognitive psychologists for within-subject designs, because the accompanying theory is typically condition-specific, giving rise to the danger that effects attributable to between-conditions interactions are attributed to within-conditions processes.

Finally, across our experiments we adhere closely to certain procedural details, such as the specific paired-associate lists presented to subjects. Such consistency permits certain between-experiments comparisons and statistical analyses thereof, but it raises certain questions as to the generality of our results. We have two reassurances to offer regarding the latter concern: (a) As cited in the discussion of Experiment 1, we have obtained the same pattern of results with different materials counterbalanced in a different fashion, and (b) when, as in Experiment 4, we use a within-lists design with our present materials, our results look typical of other within-lists results in the literature.

Experiment 1

Experiment 1 was designed to determine the relative sizes of the generation effect in free and cued recall. Encoding task (reading or generating) and type of test (free or cued recall) were manipulated between subjects. A further variable, whether the response word was a first or third associate (Palermo & Jenkins, 1964) of the stimulus, was manipulated within-subjects. This variable was included to determine if, as discussed earlier, cued recall is more sensitive to relational factors than is free recall. If it is, there should be an advantage of first over third associates in cued recall, and a lesser advantage, if any, of first associates over third associates in free recall.

Method

Subjects. The subjects were 80 introductory psychology students at the University of California, Los Angeles. They participated in the experiment in partial fulfillment of a research participation class requirement.

Design. A 2 × 2 × 2 mixed factorial design was used with encoding task (read vs. generate) and type of test (free vs. cued recall) as between-subjects factors, and associative strength (first vs. third associate) as a within-subjects factor.

Materials. Two alternative lists composed of 14 word pairs each served as the to-be-remembered materials. The 14 stimulus words in each list were the same words presented in the same order, but if a given stimulus word was paired with its first associate (Palermo & Jenkins, 1964) on one list, it was paired with its third associate on the other list. These materials are presented in the Appendix. In one list, the odd-numbered stimulus words (in terms of list position) were paired with their first associates and the even-numbered words were paired with their third associates; in the other list, the odd-numbered pairs were third associates and the even-numbered pairs were first associates. Each of the two alternative lists was presented to half of the subjects. Across subjects, therefore, assignment of first and third associates to list position and stimulus word was counterbalanced.

For each list, a 14-page booklet was constructed, with one pair typed on each page. For the read group, both words were printed on the page. For the generate group, the stimulus word was printed on the page and was followed by a fragment of the response word. Vowels were deleted from the response words to construct the fragments. Dashes represented missing letters. For both read and generate subjects there was a straight line 1 in. below the word pair. This space was provided for a written response.

The cued-recall test was constructed by randomly reordering the stimulus words and providing a space to write the response word.

Procedure. Twenty subjects, in groups of three to five, were tested in each of the encoding-task/type-of-test conditions. To increase the likelihood of random assignment of subjects to the read/generate conditions (by avoiding potential differences between successive groups of subjects), subjects in a given experimental session were randomly assigned to type of encoding task (read or generate). All of the subjects in a given experimental session was randomly assigned to type of test (free or cued). Subjects in the read and generate conditions were given different instructions (without the other subjects being present). The generate subjects were told they would be given a first word and some letter cues of a second related word. They were to figure out the second word and then write both words down. They were told that their memory would be tested, but the nature of the test was not specified. The read subjects were told they were to read a pair of related words and then write both words down. They were also told that their memory would be tested, but the nature of the test was not specified. After being given an example, both groups were told there would be 14 pairs on the list and that there would be 10 s to perform the assigned task on each pair. They were instructed in how to use the booklets and informed that the experimenter would pace them through the booklet by saying "turn" every 10 s.

After the study phase was completed, the subjects were asked to put their names on the booklets and the booklets were collected. Subjects in the free-recall groups were given a word-search puzzle for 5 min, whereas subjects in the cued-recall groups were given the same word-search puzzle and an additional word-search puzzle for a total of 20 min. The cued-recall test was delayed to avoid ceiling effects. Following the word-search task, subjects in the free-recall group were given blank sheets and asked to recall the response-word members of the pairs of words they had studied. Subjects in the cued recall group were given the cued-recall sheet and asked to recall the response word corresponding to each stimulus word. Both test groups were given 3 min for recall.

Results

In the present experiment and the experiments reported later, generation errors at the time of study were negligible. (No more than 4 out of the 20 generation subjects in each test condition made any generation errors at all, and those who did tended to make a single error on one of the 14 pairs.)

The mean proportions of correct responses in free recall and cued recall are shown in Table 1 for first and third associates as a function of encoding task. The first and last of the studied pairs were omitted from the analysis (they were considered primacy and recency buffers), so the proportions are based on six first and six third associates in the list.

As expected, the main effects of encoding task and type of test were significant, $F(1, 76) = 16.44$, $MS_e = 1.90$, $p < .0005$, and $F(1, 76) = 108.72$, $MS_e = 1.90$, $p < .0001$, respectively.

Table 1
Proportion of Response Words Recalled in Experiment 1 as a Function of Encoding Task and Level of Stimulus–Response Association

	First associates		Third associates	
Final test	Read	Generate	Read	Generate
Free recall[a]	.30	.39	.41	.47
Cued recall[b]	.82	.90	.50	.85

[a] Retention interval = 5 min. [b] Retention interval = 20 min.

The finding of primary interest is that the generation effect is larger in cued recall (.88 vs. .66) than it is in free recall (.43 vs. .36), which results in a significant interaction between encoding task and type of test, $F(1, 76) = 3.87$, $MS_e = 1.90$, $p < .05$. A planned comparison revealed that the generation effect was significant in cued recall, $F(1, 76) = 18.15$, $MS_e = 1.90$, $p < .0005$, but failed to reach significance in free recall, $F(1, 76) = 2.17$, $MS_e = 1.90$, $.10 < p < .15$.

A second result of interest is the significant interaction between type of test and associative strength, $F(1, 76) = 25.62$, $MS_e = 1.03$, $p < .0005$. Third associates were better recalled than first associates in free recall (.44 vs. .35), but the reverse was true in cued recall (.68 vs. .86). Planned comparisons revealed that both differences were statistically significant, $F(1, 76) = 5.02$, $MS_e = 1.03$, $p < .05$, and $F(1, 76) = 22.76$, $MS_e = 1.03$, $p < .0005$, respectively.

Finally, the three-way interaction was significant, $F(1, 76) = 7.6$, $MS_e = 1.03$, $p < .01$, as was the Associative Strength × Encoding Task interaction, $F(1, 76) = 4.61$, $MS_e = 1.03$, $p < .05$. No special theoretical significance is ascribed to these interactions because they may be due to a ceiling effect in the cued recall of first associates in the generate group.

Discussion

Can an item-specific explanation of the generation effect, particularly the lexical activation hypothesis, explain the interaction of encoding task and type of test obtained in Experiment 1? The answer to that question, we think, is no, but before arguing that position in some detail, we want to report the results of an experiment that was designed to address legitimate concerns that might be voiced with regard to the larger generation effect obtained in cued as opposed to free recall in Experiment 1.

Experiment 2

One concern is the generality of the materials used in Experiment 1. Although this question will not be directly addressed in this report, Hirshman (1987) has shown that the pattern of means reported in Experiment 1 recurs with a completely different set of materials. In Hirshman's experiment a different set of word pairs was used and the word-pairs were structured so that, across subjects, the to-be-recalled response terms were the same items in the pairs of high and low associative strength. Further, both the free-recall and cued-recall groups were tested at a 5-min retention interval. In this experiment the generation effect in cued recall was 9% larger than the generation effect in free recall. Thus the Type of Test × Encoding Task interaction found in Experiment 1 is not due to the peculiarities of the materials in the Appendix or to the different retention intervals used in the free-recall and cued-recall conditions.

Hirshman's (1987) results do not, however, remove questions concerning the effects of the different performance levels of free recall and cued recall found in Experiment 1. If the size of the generation effect were to be sensitive to the level of performance, the Type of Test × Encoding Task interaction found in Experiment 1 might be due to the different performance levels of free recall and cued recall. Experiment 2 was designed to address this issue by lengthening the retention interval for cued recall to 48 hr. This change was introduced to reduce the level of performance of the cued-recall group to the level of performance of the free-recall group.

Method

Experiment 2 duplicated all aspects of Experiment 1 with two exceptions. First, in Experiment 2 the cued-recall subjects were dismissed immediately after the to-be-remembered list was presented. They returned 48 hr later, at which time their cued recall of responses was tested. Second, 32 subjects received cued-recall tests and 40 subjects received free-recall tests in Experiment 2.

Results

The results of Experiment 2 are presented in Table 2. The primary finding is that the generation effect is larger in cued recall (.59 vs. .39) than in free recall (.40 vs. 34), which results in a significant interaction of type of test and encoding task, $F(1, 68) = 4.04$, $MS_e = 1.87$, $p < .05$. Planned comparisons revealed that the generation effect was significant in cued recall, $F(1, 68) = 13.33$, $MS_e = 1.87$, $p < .001$, but failed to reach significance in free recall, $F(1, 68) = 1.12$, $MS_e = 1.87$, $p > .25$. As in Experiment 1, type of test interacted with associative strength, $F(1, 68) = 23.44$, $MS_e = 1.87$, $p < .001$. Planned comparisons indicated that third associates were better free recalled than first associates, $F(1, 68) = 10.13$, $MS_e = 1.87$, $p < .005$, but the reverse was true in cued recall, $F(1, 68) = 13.32$, $MS_e = 1.87$, $p < .001$. The only other significant effects were the main effects of type of test and encoding task. Cued recall performance remained superior to free recall performance (.49 vs. .37), $F(1, 68) = 9.96$, $MS_e = 1.87$, $p < .01$, and generated items were better recalled than read items (.50 vs. .36), $F(1, 68) = 11.75$, $MS_e = 1.87$, $p < .001$. One last pattern of note is that the numerical advantage of generate

Table 2
Proportion of Response Words Recalled in Experiment 2 as a Function of Encoding Task and Level of Stimulus–Response Association

	First associates		Third associates	
Final test	Read	Generate	Read	Generate
Free recall	.31	.27	.38	.53
Cued recall	.48	.71	.29	.48

over read in free recall occurred only for third associates. This point is returned to in the General Discussion.

One purpose of Experiment 2 was to replicate the free-recall conditions of Experiment 1 in order to ascertain whether the marginally nonsignificant advantage of generate over read found in Experiment 1 should be interpreted as a real difference. Toward that end, a single analysis of variance was carried out on the free recall data of Experiments 1 and 2 with experiment and encoding task as between-subjects variables and associative strength as a within-subjects variable. The results of this analysis reveal a significant three-way interaction, $F(1, 76) = 3.95$, $MS_e = 1.15$, $p < .05$, and significant main effects of encoding task, $F(1, 76) = 5.19$, $MS_e = 1.15$, $p < .05$, that is, a generation effect in free recall, and associative strength, $F(1, 76) = 20.17$, $MS_e = 1.15$, $p < .001$.

To investigate the three-way interaction, the interaction of associative strength and encoding task was examined separately for the free recall conditions of Experiments 1 and 2. Planned comparisons revealed that encoding task interacted with associative strength in Experiment 2, $F(1, 76) = 5.73$, $MS_e = 1.15$, $p < .01$, but not in Experiment 1 ($F < 1$). In Experiment 2 there was a numerical advantage of generate over read only on third associates. In Experiment 1 this numerical advantage held for both first and third associates. The main effect of encoding task indicated that generate items were better free recalled than read items (.42 vs. .35), and the effect of associative strength indicated that third associates were better free recalled than first associates (.44 vs. .32). There were no other significant effects.

The results of the preceding analysis indicate that the small (7%) generation effect found in free recall is reliable. To replicate and extend this conclusion, the free-recall conditions of Experiment 1 were again replicated with one minor difference. In this replication subjects were told before the study phase that they were going to be asked to free recall the response words. The results mirrored those of the previous free-recall conditions. There was again a small advantage in free recall of generate response words over read response words (.48 vs. .40 on first associates and .56 vs. .43 on third associates). This effect was marginally nonsignificant, $F(1, 38) = 3.05$, $MS_e = 2.35$, $p = .08$. An additional analysis compared the results of the replication experiment just mentioned and the free-recall conditions of Experiments 1 and 2. Type of experiment and encoding task were between-subjects variables in this analysis. Associative strength was a within-subjects variable. Response recall in the replication was superior to response recall in Experiments 1 and 2 (.47 vs. .38), $F(1, 116) = 9.33$, $MS_e = 1.53$, $p < .005$, but type of experiment did not interact with encoding task ($F < 1$). Furthermore, across the three experiments the effect of generation was significant (.45 vs. .37), $F(2, 116) = 8.45$, $MS_e = 1.53$, $p < .005$, and third associates were better recalled than first associates (.46 vs. .35), $F(1, 116) = 12.66$, $MS_e = 1.30$, $p < .001$. There were no other significant effects. The combined results of the free-recall conditions of Experiments 1 and 2, and the replication just mentioned provide strong evidence that there are small (6%–10%) generation effects in free recall when a between-subjects design is used.

Discussion

The results of Experiments 1 and 2 appear to rule out any one-factor item-specific theory, such as the lexical activation hypothesis, as a sufficient explanation of the generation effect. That such a conclusion is warranted can be illustrated in the following fashion. Assume, as shown in Figure 1, that as a consequence of a given read or generate study trial a certain number of features are activated in memory. Across items and subjects there will be a distribution of activated features in both cases owing to item variability and subject variability (both between subjects and within a given subject across trials as level of attention and effort varies). The upper panel in Figure 1 illustrates the case in which the two distributions are presumed to have equal variance (there is no a priori reason to assume otherwise). The lower panel illustrates the case in which the variance of the generate distribution is assumed to be smaller than the variance of the read distribution, a relation that turns out to be required by the results of Experiment 1. In either case, of course, the mean of the generate distribution is assumed to be higher than the mean of the read distribution, which is required by any one-factor item-specific explanation of the generation effect.

The final assumption incorporated in Figure 1 is that later cued or free recall is successful when the number of features activated at study exceeds a criterion, where the criterion number of features for cued recall, B_{cr}, is a smaller number than the criterion number for free recall, B_{fr}. (Because the

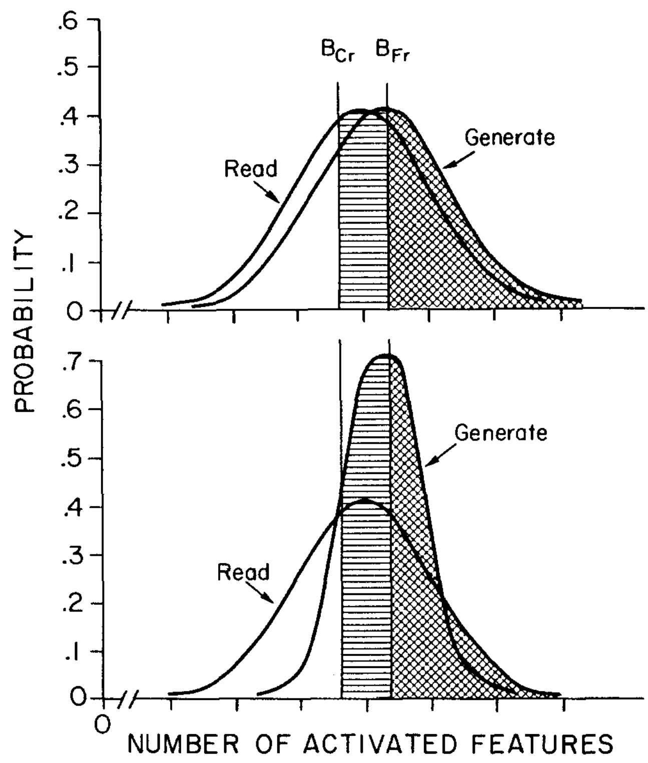

Figure 1. Theoretical distribution of the number of lexical features activated by a read or generate encoding task.

cued-recall test is an easier test than the free-recall test, fewer activated features are necessary for recall to succeed.) The expected proportion correct, then, is simply the area under the appropriate distribution to the right of the relevant criterion. The upper panel of Figure 1 illustrates the case where the standard deviations of the two distributions are assumed to be one, and the two criteria are placed to fit the observed proportions of cued and free recall in the read condition.

In terms of the representation in Figure 1, the results of Experiment 1 determine the z scores for the free-recall and cued-recall criteria for the read and generate distributions. For the sake of simplicity, performance in Experiment 1 is shown averaged over first and third associates. (The following demonstration can also be given separately for first and third associates across the pattern of results reported in this article.) The z scores are,

For the read distribution,

$$B_{cr}: -0.41$$

$$B_{fr}: +0.36,$$

and, for the generate distribution,

$$B_{cr}: -1.18$$

$$B_{fr}: +0.18.$$

These imply

$$B_{fr} - B_{cr} = 0.77\sigma_r \quad (1)$$

for the read distribution, and

$$B_{fr} - B_{cr} = 1.36\sigma_g \quad (2)$$

for the generate distribution, where σ_r = the standard deviation of the read distribution, and σ_g = the standard deviation of the generate distribution.

Equal variances. If we assume that the standard deviations of the two distributions are equal (as in the upper panel of Figure 1), then there are no positions of the criteria (B_{cr}, B_{fr}) that accommodate the results of Experiment 1. The criteria cannot be both 0.77 and 1.36 standard deviations apart.

Unequal variances. If the standard deviations of the read and generate distributions are permitted to be unequal, then there are values of the criteria that accommodate the results of Experiment 1. Equations 1 and 2 require that the criterion values be characterized by the following relation between the standard deviations of the two distributions:

$$1.36\sigma_g = 0.77\sigma_r, \quad (3)$$

which can be rewritten as

$$\sigma_g = 0.57\sigma_r. \quad (4)$$

Thus, to account for the results of Experiment 1, the generate distribution needs to have substantially smaller variance than the read distribution. The case where σ_r is assumed to be one is shown in the bottom panel of Figure 2.

Any argument as to whether it is plausible that the generate condition should yield a distribution of features that has a smaller variance than the read distribution is academic because the solution in the lower panel of Figure 1 is inconsistent with the results of the 48-hr cued recall in Experiment 2. As the cued-recall test is delayed and, hence, becomes more difficult, one needs to assume that B_{cr} moves to the right (or alternatively, that the distributions move to the left). As that happens the advantage of generate over read will decrease because more area under the generate distribution will move to the left of the criterion than will area under the read distribution. (In fact, the unequal variance case predicts an advantage of read over generate if a test were to be made difficult enough.) Once the overall level of cued recall is lowered to the approximate level of free recall in Experiment 1, the size of the generation effect should be about the same according to the representation in Figure 1. The results of Experiment 2 indicate, however, that the generation effect in cued recall is significantly larger than the generation effect in free recall even when performance levels in free recall and cued recall are approximately the same.

What is important about the foregoing demonstration is not the demonstration that a one-factor theory in the form of an item-specific difference in the number of features activated is untenable, but that any one-factor item-specific theory is untenable. That is, any theory that accounts for the generation effect solely in terms of the assertion that the generate condition yields more of something vis-a-vis the to-be-recalled response term of a studied paired associate is untenable. It does not matter whether the "something" is lexical features, number of stored operations or whatever.

Thus, in addition to what item-specific advantages the generate condition might have, it appears necessary to assume that the act of generation also strengthens something else more than does the act of reading, and the natural candidate for that "something else" appears to be the stimulus–response relation (or bond) between the stimulus and response words. According to the two-process theory discussed in the introduction, we would expect both relational and item-specific factors to contribute to the generation effect in cued recall, whereas only item-specific factors contribute to the generation effect in free recall. The finding in Experiments 1 and 2 that third associates are better free recalled than first associates supports the assumption that a stronger relation between the items in a pair does not increase, and can even inhibit the free recall of items in that pair.

The foregoing framework predicts that it should be possible to remove the generation effect in free recall without removing the generation effect in cued recall. Experiment 3 was designed to test this prediction.

Experiment 3

In Experiments 1 and 2 subjects were required in both the read and generate conditions to write down both members of the given stimulus–response pair. In Experiment 3 the procedure was modified by requiring subjects to write down only the response word after the act of reading or generating. Such a procedure should focus attention on the response word, which, we assume, will facilitate the read condition more than it will the generate condition in free recall. The argument is that in the generate condition, the response word is generated just prior to its being written down, so that the act of writing that word is of little additional benefit. In the read condition, however, the act of writing down the response word should serve to enhance substantially the activation of response-word-specific features in memory.

This procedure, however, should not remove the generation effect in cued recall because the generation effect in cued recall is dependent on both relational and item-specific factors.

Method

Subjects. The subjects were 156 introductory psychology students at the University of California, Los Angeles. They participated in partial fulfillment of a research-participation class requirement. Seventy-six subjects were in the free-recall group, and 80 subjects were in the cued-recall group.

Design and materials. The design and materials were the same as those used in Experiment 1.

Procedure. The procedure duplicated that of Experiment 1 except that instead of writing down both words, following the read or generate task, subjects only wrote down the response word.

Results

The results of Experiment 3 are shown in Table 3. As expected the effects of type of test and encoding task are significant, $F(1, 152) = 218.56$, $MS_e = 1.83$, $p < .001$, and $F(1, 152) = 8.48$, $MS_e = 1.83$, $p < .01$, respectively.

The result of primary interest is that the generation effect occurred in cued recall but not free recall. The interaction of Encoding Task × Type of Test was significant, $F(1, 152) = 14.22$, $MS_e = 1.83$, $p < .001$. Planned comparisons revealed a significant effect of generation in cued recall, $F(1, 152) = 22.92$, $MS_e = 1.83$, $p < .0001$, but no effect of generation in free recall ($F < 1$). In fact, there is a slight numerical advantage of the read group over the generate group in free recall (.45 vs. .43). As in Experiments 1 and 2, type of test interacted with associative strength, $F(1, 152) = 29.97$, $MS_e = 1.04$, $p < .001$. Third associates were better free recalled than first associates, $F(1, 152) = 5.64$, $MS_e = 1.04$, $p < .05$, but the reverse was true in cued recall, $F(1, 152) = 29.21$, $MS_e = 1.04$, $p < .001$.

Finally, there was a significant interaction between encoding task and associative strength, $F(1, 152) = 12.69$, $MS_e = 1.04$, $p < .001$. To investigate this interaction further, the Encoding Task × Associative Strength interaction was examined separately for free recall and cued recall. There was a marginally nonsignificant interaction of encoding task and associative strength in cued recall, $F(1, 152) = 3.43$, $MS_e = 1.04$, $p = .06$. No importance is ascribed to this interaction because it may be due to a ceiling effect on the cued recall of generated first associates. The Encoding Task × Associative Strength interaction is significant in free recall, $F(1, 152) = 10.04$, $MS_e = 1.04$, $p < .01$. Planned comparisons indicate that first associates in the read condition are better free recalled than first associates in the generate condition, $F(1, 152) = 7.59$, $MS_e = 1.08$, $p < .01$, but this is not true for third associates. For third associates, there is a numerical generation advantage comparable to those found in the free-recall conditions of Experiments 1 and 2. This pattern indicates that the manipulation used in Experiment 3 eliminated the generation effect in free recall primarily through its influence on first associates.

It is important to note that the free-recall conditions on Experiment 3 used twice as many subjects as the free-recall conditions of Experiment 1 or 2. This ensures the stability of the means in Experiment 3 and makes the results of Experiment 3 directly comparable to the results of the combined analysis performed on Experiments 1 and 2. This comparison clearly indicates that there is a small generation effect in free recall, but that the manipulation used in Experiment 3 removes this effect.

Table 3
Proportion of Response Words Recalled in Experiment 3 as a Function of Encoding Task and Level of Stimulus–Response Association

	First associates		Third associates	
Final test	Read	Generate	Read	Generate
Free recall	.47	.36	.44	.51
Cued recall	.84	.96	.63	.86

Discussion

The results of Experiment 3 are consistent with the two-factor theory of the generation effect outlined earlier. In Experiment 3 a minor procedural change eliminated the advantage of generate over read found in free recall in Experiments 1 and 2 but did not remove the effect of generation in cued recall.

Between-subjects versus within-subjects designs. A major problem with the foregoing account is that it is based on an empirical finding that conflicts with most of the published literature on the generation effect. Slamecka and Graf (1978), McFarland et al. (1980), Johnson et al. (1981), and Gardiner and Arthurs (1982) have reported large, reliable generation effects in free recall with similar materials and generation tasks to those we used. The primary difference between our experiments and those reported in the foregoing articles is our use of a between-subjects design. With one exception, every free recall experiment in the other articles used a within-subjects design. Interestingly, in the one experiment that used a between-subjects design, McFarland et al. (1980, Experiment 2B) found the effect of generation in free recall is approximately half the size of the effect of generation in their Experiment 1, a within-subjects design.

Pair-specific versus whole-list effects of generation. In the introduction we emphasized the importance of distinguishing between the pair-specific and whole-list effects of generation. Although most theoretical accounts focus on pair-specific effects (and generally ignore whole-list effects), whole-list effects do occur and are confounded with pair-specific effects. Specifically, in within-subjects designs, read and generate items compete for attentional resources and interfere with each other at output. Ignoring such factors leads to a systematic overestimation of the pair-specific effects of generation per se. In particular, the use of within-subjects designs can overestimate the size of the generation effect in free recall, and we contend this overestimation has obscured the critical interaction between type of test and encoding task reported in this article.

Experiment 4 was designed to test the foregoing speculations and duplicates Experiment 1 with one important exception. Encoding task (read vs. generate) was a within-subjects variable in Experiment 4. The contrast between Experiment 4 and Experiments 1 and 2 provides a direct test of our speculations that (a) the generation effect is larger in within-subjects designs than in between-subjects designs; (b) the critical difference between Experiments 1 and 2 in this article and previous free-recall experiments in the literature is the use of a between-subjects design; and (c) by confusing the pair-specific and whole-list effects of generation, previous authors have systematically overestimated the average generation effect in free recall and have concealed a theoretically important Type of Test × Encoding Task interaction.

Experiment 4

Method

Subjects. The subjects were 40 introductory psychology students at the University of California, Los Angeles. They participated in the experiment in partial fulfillment of a research-participation class requirement.

Design and materials. A 2 × 2 mixed factorial design was used with type of test (free recall vs. cued recall) as a between-subjects factor and encoding task (read vs. generate) as a within-subjects factor. In Experiment 1 there was a read and generate version of each of the two lists presented to subjects. Experiment 4 manipulated encoding task within-subjects by exchanging items between the read and generate version of a given list. Specifically, the even-numbered items (in terms of list position) were switched between the read and generate versions of a given list. This produced a total of four lists. In two lists read items were in the odd-numbered list positions and generate items occupied the even-numbered list positions; for the other two lists the opposite was true. Furthermore, in two of the lists read items were third associates of their stimulus terms, whereas generate items were first associates of their stimulus terms. In the other two lists that assignment was reversed. Subjects were randomly assigned to one of the four lists. Across subjects, therefore, assignment of read versus generate task to list position, item type, and associative relation of that item were counterbalanced. Note that unlike the previous experiments, associative strength is not factorially crossed with encoding task and type of test in Experiment 4.

Procedure. The procedure of Experiment 4 duplicated that of Experiment 1 with three exceptions. First, instead of being assigned to a read or generate group, subjects were randomly assigned to one of the four lists. Second, they were given examples of both read and generate encoding tasks prior to the presentation of the list. Third, only 20 subjects (instead of 40) received free-recall tests and only 20 subjects (instead of 40) received cued-recall tests in Experiment 4. As in Experiment 1 the retention interval was 5 min for the free-recall condition and 20 min for the cued-recall condition.

Results and Discussion

The results of Experiment 4 are presented in Table 4. As expected, the effects of type of test and encoding task were significant, $F(1, 38) = 45.43$, $MS_e = 2.53$, $p < .0001$, and $F(1, 38) = 41.19$, $MS_e = 1.84$, $p < .0001$, respectively. Unlike the results of Experiments 1, 2, and 3, the Encoding Task × Type of Test interaction was not significant ($F < 1$). The difference between the generate and read means is still larger in cued recall (.38) than in free recall (.28). Planned comparisons indicated the effect of generation was significant in both free recall and cued recall, $F(1, 38) = 14.74$, $MS_e = 1.84$, $p < .001$, and $F(1,38) = 27.42$, $MS_e = 1.84$, $p < .0001$, respectively.

Table 4
Proportion of Response Words Recalled in Experiment 4 as a Function of Encoding Task and Type of Test

Test	Read	Generate
Free recall	.13	.41
Cued recall	.48	.86

These results clarify several important points. First, the generation effect is larger in within-subjects designs than in between-subjects designs. In free recall the difference between the generate and read means is .28 in within-subjects designs and .07 in between-subjects designs. In cued recall this same difference is .38 in within-subjects designs and .22 in between-subjects designs. Second, the materials and procedures we used produced a large, reliable generation effect in free recall when a within-subjects design is used. This result is consistent with the results reported by other investigators.

A comparison of the results of Experiments 1 and 4 illuminates the difference between the two types of designs. The effect of generation was larger in Experiment 4 (within subjects) than in Experiment 1 (between subjects) because the means of the read groups were much lower in Experiment 4 than in Experiment 1. This pattern occurred in free recall and, to a lesser extent, in cued recall. The free-recall mean in the read condition fell by .23, whereas the read-cued-recall mean fell by .18. These results are even more dramatic when the performance of the read group in the within-subjects design is represented as a percentage of the read group performance in the between-subjects design; in free recall the within-subjects mean is 33% of the between-subjects mean (.13 vs. .36). In cued recall, for the read condition, the within-subjects mean is 63% of the between-subjects mean (.48 vs. .66). This pattern indicates that the additional effects of generation that emerge in the within-subjects design operate to lessen performance on the read items. This outcome is consistent with Begg and Snider's (1987) claim that the presence of generate items in a list entails cursory processing of read items in that list, as well as with our claims that generate items compete with read items at encoding and output.

One peculiarity of the design of Experiment 4 remains to be discussed. In two of the lists used in Experiment 4, read items were third associates and generate items were first associates. In the other two lists this assignment was reversed: Generate items were third associates and read items were first associates. Unlike previous experiments, the factors of encoding task and associative strength are not factorially crossed in Experiment 4. To investigate whether this peculiar list arrangement had any effect, the results of Experiment 4 were reanalyzed with list type (generate-first associate, read-third associate vs. read-first associate, generate-third associate) as a third factor. The main effect of the factor of list type was not significant ($F < 1$) and list type did not interact with either encoding task or type of test. The three-way interaction was also not significant ($F < 1$).

General Discussion

The strongest implication of our results is that a one-factor theory is not sufficient to explain the pair-specific effects of generation. The results are also consistent with a two-factor theory that we consider to be plausible. In the remainder of this section we discuss (a) those presumed two processes and their interaction with type of test and level of stimulus–response associative strength, (b) some alternative interpretations, and (c) some loose ends and remaining issues.

The empirical facts, as shown in this report, are that the generation effect is much smaller in free recall than in cued recall, and that first associates, although clearly better recalled in cued recall, are more poorly recalled in free recall. The small, but reliable, effects of generation in free recall can be eliminated by the kind of minor procedural change introduced in Experiment 3 without removing the generation effect in cued recall.

Those empirical facts are accounted for nicely by a theory that assumes generating is superior to reading as a study condition in two respects: It is better at activating features of the response term in memory, and it strengthens the stimulus–response relation in memory. In addition, we assume the following. (a) Cued recall is sensitive primarily to stimulus–response strength but is facilitated by response activation as well. Such an assumption is consistent with the classic view that paired-associate learning involves response learning as well as stimulus–response association or "hookup" (cf. Underwood & Shulz, 1960). Consequently, the generation effect in cued recall is due to differences in stimulus–response strength and response activation between read and generated items. (b) Free recall of responses is sensitive primarily to response strength or activation and is inhibited rather than facilitated by stimulus–response strength. Such an assumption is based in part on prior evidence, cited earlier in this article that increased unitization (association) reduces access to components of the unit, and in part on the fact that the present data on free recall of first and third associates seem to demand such an assumption. Consequently, the generation effect in free recall of responses is due to response strength or activation, and is inhibited rather than facilitated by stimulus–response strength.

Given the foregoing assumptions, all of the following aspects of the current results are interpretable. (a) Free recall in both the read and generate conditions is better for third associates than for first associates; the opposite is true for cued recall. Such a result follows because stimulus–response strength enhances cued recall but inhibits free recall. (b) The generation effect is larger in cued than in free recall. Both advantages of generating over reading (stimulus–response strengthening, response activation) will contribute to the generation effect in cued recall. In free recall, however, only response activation contributes to the generation effect. (c) When the encoding task is changed by having the read and generate subjects write down only the response term, the generation effect in free recall is eliminated without removing the generation effect in cued recall. That minor procedural change should, according to the arguments presented earlier, increase the response activation more in the read condition than in the generate condition. According to the present theory this should remove the generation effect in free recall because this effect depends solely on response activation. However, the generation effect should remain in cued recall because this effect depends on the generate condition's stimulus–response relational advantage as well as its response-activation advantage. These are the results actually found in Experiment 3.

Finally, as noted earlier, type of experimental design is an important factor affecting the size of the generation effect. In within-subjects designs, whole-list factors decrease the performance of the read group so that the generation effect is larger in within-subject designs than in between-subject designs. If one is interested in studying whole-list interactions of conditions on attentional and output processes, then a within-list procedure may be appropriate. If one's theorizing, on the other hand, focuses on the pair-specific effects of generation per se, then the within-list procedure confounds whole-list processes with the pair-specific processes of interest. This can camouflage theoretically important pair-specific effects of generation such as the Encoding Task × Type of Test interaction found in Experiments 1, 2, and 3.

Alternative Interpretations of the Present Results

There are two alternatives to the foregoing theory that merit comment. The first theory is an alternative view of the effects of the relational and response-activation factors in the generation effect. The second view presents an additional or alternative factor to the two we have presented.

First, it could be argued, on the basis of the pattern of generation effects obtained in the present experiments, that free recall is independent of the relational factor and cued recall is independent of the response-specific factor, but that the relational factor is more potent than the response-specific factor, which yields a larger effect of generation in cued recall than in free recall. That the relational factor is potent, and specific to the particular relation strengthened at study, is indicated by the recent work of Rabinowitz and Craik (1986), who found that generating a response based on a semantic rule did not enhance cued recall over that of a read group when a rhyme cue was used at test. Thus, the response activation benefits of generation do not seem sufficient to create a generation effect in cued recall if the relational factor is nullified by the kind of change used by Rabinowitz and Craik.

Such an alternative interpretation does not seem viable, however, for at least two reasons. First, the overall advantage in free recall of third associates over first associates shows that free recall is dependent on stimulus–response strength. The negative nature of that dependency argues for an inhibitory relation between stimulus–response strength and availability of the response term in free recall. Second, Gardiner and Hampton (1985) have demonstrated substantial generation effects in free recall using a transposition rule, where it seems very unlikely that a relational factor could mediate the effect. Such a result argues against the idea that increased response activation is a minor advantage of generation as it might appear to be in our experiments.

Nairne et al. (1985) presented a second alternative view. Whereas we have emphasized response-specific activation and the strengthening of the relation from the stimulus term to the response term, Nairne et al., on the basis of the failure to find a generation effect with low-frequency words, argued that generation strengthens the relation from a response to its associates. This view does not seem fully compatible with our data. Because stimuli paired with first associates are much more likely to be high associates of their response terms than are stimuli paired with third associates, we would expect the generation effect in cued recall to be larger when first-associate stimuli are used as cues than when the third associate stimuli are used as cues. As will be discussed next, there is only marginal support for this position.

Remaining Issues and Loose Ends

An important and interesting issue is whether the effect of generation differs as a function of associative level. It seems reasonable to expect that both advantages of generation would be larger for third associates than for first associates. The more difficult act of generation in the third associate case should lead to greater response activation, and the potential for increasing the stimulus–response association over the preexperimental baseline would seem greater for third associates as well.

Unfortunately, our data are ambiguous with respect to that issue. Looking only at the free-recall results of Experiments 1, 2, and 3 and the replication presented in the Discussion of Experiment 2, one could argue that the response activation effect of generation is greater for third associates. Across those four experiments, the average proportion of first associates recalled was .37 for read and .37 for generate, whereas the average proportion of third associates recalled was .42 for read and .52 for generate. Because only the response-activation factor contributes positively to free recall, and because the relational strengthening, which acts negatively, is, as will be argued next, equal across first and third associates, such a pattern of results suggests that the effect of generation on response activation is greater for third associates than for first associates.

There does not seem to be any good evidence, however, that the effect of generation in stimulus–response strengthening is greater for third associates. Assuming that free recall would decline to zero (or close to zero) by 48 hr, we can assume that response activation in 48-hr cued-recall (Experiment 2) is zero and that the results of this experiment provide us a pure measure of relational strength. These results indicate that the act of generation increments relational strength as much for first associates as for third associates. Although Experiments 1 and 3 suggest that the generation effect is larger for third associates in cued recall, they are not definitive on this point, in part because of the ceiling effect on the cued recall of first associates, and in part because the presumed greater activation in the generate condition for third associates contributes an unknown positive amount to the cued recall of third associates in the generate condition.

Finally, it is clear that read and generate study conditions differ in an important respect not addressed herein. The *data-driven* consequences of reading a response word versus generating that word appear to differ markedly. On nontraditional measures of memory that are particularly sensitive to data-driven processing, such as perceptual identification and word-fragment completion, there is the opposite effect of generation. Jacoby (1983) has demonstrated a stunning crossover interaction of study condition (read vs. generate) and type of final test (recognition vs. perceptual identification) characterized by an advantage of read over generate on the perceptual identification task. Blaxton (reported in Roediger & Blaxton, 1987) has obtained a similar crossover interaction with free recall and word-fragment completion as the final tasks.

Such interactions support the distinction between data-driven (or stimulus-driven) and conceptually driven processing (see Roediger & Blaxton, 1987). A read condition serves to prime later perceptual identification or word-fragment completion better than does a generate condition because the target word is presented at study with its physical features intact, rather than, as in the generation case, with certain letters replaced by dashes.

Our analysis has focused on differences in the conceptually driven aspects of reading and generating. Such differences are the pertinent differences for analyses of later recall (for a discussion of what types of measures reflect what types of processing, see Richardson-Klavehn and Bjork, in press). Viewed from a broader standpoint, however, the two differences we identify as responsible for the generation effect do not exhaust the processing differences between reading and generating.

References

Anderson, J. R. (1972) FRAN: A simulation model of free recall. In G. H. Bower (Ed.), *The psychology of learning and motivation* (Vol. 5, pp. 315–378). New York: Academic Press.

Begg, I., & Snider, A. (1987). The generation effect: Evidence for generalized inhibition. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 13*, 553–563.

Donaldson, W., & Bass, M. (1980). Relational information and memory for problem solutions. *Journal of Verbal Learning and Verbal Behavior, 19*, 26–35.

Gardiner, J., & Arthurs, F. (1982). Encoding context and the generation effect in multi-trial free recall learning. *Canadian Journal of Psychology, 36*, 527–531.

Gardiner, J., & Hampton, J. (1985). Semantic memory and the generation effect: Some tests of the lexical activation hypothesis. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 11*, 732–741.

Gardiner, J., & Rowley, J. (1984). A generation effect with numbers rather than words. *Memory & Cognition, 12*, 443–445.

Ghatala, E. (1983). When does internal generation facilitate memory for sentences. *American Journal of Psychology, 96*, 78–83.

Glisky, E., & Rabinowitz, J. (1985). Enhancing the generation effect through repetition of operations. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 11*, 193–205.

Graf, P. (1980). Two consequences of generating: Increased inter and intra word organization of sentences. *Journal of Verbal Learning and Verbal Behavior, 19*, 316–327.

Graf, P. (1982). The memorial consequences of generation and transformation. *Journal of Verbal Learning and Verbal Behavior, 21*, 539–548.

Hayes-Roth, B. (1977). Evolution of cognitive structures and processes. *Psychological Review, 84*, 260–278.

Hirshman, E. (1987). *The unitization effect: Paradoxical effects of semantic relatedness.* Unpublished doctoral dissertation, University of California, Los Angeles.

Hunt, R., & Einstein, G. (1981). Relational and item-specific information in memory. *Journal of Verbal Learning and Verbal Behavior, 20*, 497–514.

Jacoby, L. (1978). On interpreting the effects of repetition: Solving a problem versus remembering a solution. *Journal of Verbal Learning and Verbal Behavior, 17*, 649–667.

Jacoby, L. (1983). Remembering the data: Analyzing interactive processes in reading. *Journal of Verbal Learning and Verbal Behavior, 22*, 485–508.

Johnson, M., Raye, C., Foley, H., & Foley, M. (1981). Cognitive operations and decision bias in reality monitoring. *American Journal of Psychology, 94*, 37–64.

McElroy, L., & Slamecka, N. (1982). Memorial consequences of generating nonwords: Implications for theories of the generation effect. *Journal of Verbal Learning and Verbal Behavior, 21*, 249–259.

McFarland, C., Frey, T., & Rhodes, D. (1980). Retrieval of internally versus externally generated words in episodic memory. *Journal of Verbal Learning and Verbal Behavior, 19*, 210–225.

Nairne, J., Pusen, C., & Widner, R. (1985). Representation in the mental lexicon: Implications for theories of the generation effect. *Memory & Cognition, 11*, 183–191.

Nairne, J., & Widner, R. (1987). Generation effects with nonwords: The role of test appropriateness. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 13*, 164–171.

Palermo, D., & Jenkins, J. (1964). *Word association norms.* Minneapolis: University of Minnesota Press.

Payne, D., Neely, J., & Burns, D. (1986). The generation effect: Further tests of the lexical activation hypothesis. *Memory & Cognition, 14*, 246–252.

Poulton, E. C. (1982). Influential companions: Effects of one strategy on another in the within-subject designs of cognitive psychology. *Psychological Bulletin, 91*, 673–690.

Rabinowitz, J., & Craik, F. I. M. (1986). Specific enhancement effects associated with word generation. *Journal of Memory and Language, 25*, 226–237.

Richardson-Klavehn, A., & Bjork, R. A. (in press). Measures of memory. *Annual Review of Psychology.*

Roediger, H. L., & Blaxton, T. A. (1987). Retrieval modes produce dissociations in memory for surface information. In D. Gorfein & R. R. Hoffman (Eds.), *Memory and cognitive processes: The Ebbinghaus centennial conference* (pp. 349–379). Hillsdale, NJ: Erlbaum.

Slamecka, N., & Fevreiski, J. (1983). The generation effect when generation fails. *Journal of Verbal Learning and Verbal Behavior, 22*, 153–163.

Slamecka, N., & Graf, P. (1978). The generation effect: delineation of a phenomenon. *Journal of Experimental Psychology: Human Learning and Memory, 14*, 592–604.

Underwood, B., & Schulz, R. (1960). *Meaningfulness and verbal learning.* Chicago: Lippincott.

Yekovich, F. R., & Manelis, L. (1980). Accessing integrated and nonintegrated propositional structures in memory. *Memory & Cognition, 8*, 133–140.

Appendix

Materials

Stimuli	First associate	Third associate
Sickness	Health	Illness
Rough	Smooth	Hard
Eating	Food	Hungry
Always	Never	Sometimes
Hammer	Nail	Tool
High	Low	Ladder
House	Home	Roof
King	Queen	Ruler
Speak	Talk	Softly
Scissors	Cut	Paper
Bed	Sleep	Pillow
Dark	Light	Room
Quickly	Fast	Run
Chair	Table	Soft

Received August 29, 1986
Revision received June 15, 1987
Accepted June 22, 1987 ■