Representation in the mental lexicon: Implications for theories of the generation effect

JAMES S. NAIRNE, CONSTANCE PUSEN, and ROBERT L. WIDNER, JR.
University of Texas, Arlington, Texas

Three experiments investigated the finding by McElroy and Slamecka (1982) that the "generation effect" (the retention advantage for self-produced over read items) is not obtained when artificial, meaningless nonwords are used as the to-be-remembered items. In Experiment 1, some subjects were asked to generate or read items that they thought were words, but, in fact, were not; no generation effect was found. In Experiment 2, subjects were taught definitions to experimenter-created items. Despite the fact that these subjects could readily retrieve each item's assigned semantic properties, no generation effect was found. Finally, Experiment 3 examined the read/generate variable as a function of an item's frequency of use in the language. Whereas medium- and high-frequency words produced large generation effects, no comparable effects were found for low-frequency words or nonwords. These results indicate that representation in the mental lexicon is a necessary, but not a sufficient, condition for the generation effect. Rather, it may be necessary to consider how related the generated item is to other potential retrieval cues in the memory system.

In a number of recent studies, a retention advantage has been demonstrated for items that are self-produced, rather than read, by subjects during the study phase of a memory experiment. For example, if one subject is asked to generate a word that rhymes with "cart" (e.g., "start") and another subject is asked only to verify that the words "cart" and "start" rhyme, then one can expect better memory for the generated rhyme (e.g., McFarland, Frey, & Rhodes, 1980; Slamecka & Graf, 1978). This advantage, which is referred to as the "generation effect," does not depend exclusively upon the use of "deep" or semantic production rules; rather, the effect remains quite robust across wide variations in generation rules (e.g., semantic, rhyme, orthographic) and methods of retention testing (e.g., cued recall, free recall, recognition). Still, although the empirical status of the basic phenomenon is relatively clear, its theoretical underpinnings remain somewhat of a mystery.

Following a recent discussion by McElroy and Slamecka (1982), we can outline two main classes of explanation: (1) those interpretations that appeal to the special involvement of lexical memory processes during generation (e.g., Donaldson & Bass, 1980; Graf, 1980; Slamecka & Graf, 1978) and (2) accounts that refer primarily to the effortful nature of the act of generation, with no special reliance upon lexical memory involvement (Griffith, 1976; McFarland et al., 1980). By lexical or semantic memory, we are referring simply to a person's preexisting knowledge about verbal information, including, in particular, words and their corresponding definitions. Thus, we can ask, does the generation advantage result from searching and then accessing an entry in the mental lexicon? Or, does generation require the subject simply to spend more time and effort, than does reading, during encoding? Presumably, producing a word demands more processing effort than reading does, and this may translate into a stronger or more accessible memory trace (Tyler, Hertel, McCallum, & Ellis, 1979). In contrast, increased "mental work" may be of little value unless it acts to integrate an existing lexical entry with other entries in an established lexical network.

McElroy and Slamecka (1982) sought to discriminate between these two alternatives by obtaining a generation effect with nonwords. In defense of a lexical activation hypothesis, they reasoned that since nonwords contain no lexical entry, one should be unable to obtain a generation effect with such stimuli because no contact with lexical memory is possible. Instead, if a generation effect with nonwords is found, they argued, then support should be given to a simple effort hypothesis. In three separate experiments using experimenter-created nonwords as stimuli, McElroy and Slamecka failed to find an advantage for the generated over the read nonword items. Since similar experimental conditions with actual words produced an effect, the authors concluded that lexical memory involvement is necessary for obtaining the generation advantage. In order for the act of generation to enhance retention, the subject must be able to access a preexisting lexical address during the generation process.

The experiments reported in the present article were designed to extend the empirical basis for McElroy and Slamecka's (1982) conclusions, thereby allowing us to fine-tune theoretical interpretations of the generation ef-
fect. In each of our experiments, subjects were asked to generate or read "nonwords" prior to a final test of retention. Our studies differed from those of McElroy and Sla- mecka, however, in the treatment of what constituted a "nonword." In Experiment 1, some subjects were asked to generate or read items that they thought were words, but, in fact, were not. Our intention was to see whether knowledge about the lexical status of an item affects the likelihood of obtaining a generation advantage. In Experiment 2, subjects were taught definitions to experimenter-created items. In this case, we were interested in whether the simple assignment of "meaning" to a nonword would be sufficient to make it more memorable when generated rather than read. Finally, Experiment 3 compared the size of the generation effect across items that differed in frequency of use in the general language. Although low- and high-frequency words are both represented in the lexicon, they do not possess the same degree of association with other information in the lexical network. Is the generation effect found for any firmly established entry in the lexicon? Or, does the size of the effect depend upon an item's "spread" in the network—that is, the number of associative connections that the item may have to other represented lexical entries?

EXPERIMENT 1

Experiment 1 had two major goals. First, we intended to replicate McElroy and Sla-mecka (1982) by examining the memorial consequences of passive reading versus active generation of nonwords. For the generation task, we used the "rhyme rule" employed by those authors in their third experiment: Subjects were required to produce a nonword, given an initial letter, that sounded similar to a cue word (e.g., PRAB F___). For the read condition, both the cue and target words were completely revealed (e.g., PRAB FRAB). As Sla-mecka and Graf (1978) demonstrated, rhyme-based generation rules—with real-word pairs—produce robust generation advantages (see, also, McFarland et al., 1980). In the present case, the empirical question asked whether subjects would be more likely to remember generated than read nonwords in a later test of free recall.

Our second goal in Experiment 1 was to examine whether knowledge about the lexical status of the items might account for the reported failure to find a generation effect with nonwords. In McElroy and Sla-mecka’s (1982) experiments, subjects were always aware that they were dealing with artificial, experimenter-created nonwords. Under such conditions, subjects may be biased against any attempt to access the lexicon during generation; consequently, they may not show any beneficial effect of such a potentially effortful search process. It is quite conceivable that the act of generation, regardless of the particular rule employed (e.g., rhyme, orthographic, or semantic), encourages the subject to search the lexicon (i.e., with whatever generation cues are provided) because, with words, it is always a lexical en-
try that needs to be produced. For "read" items, lexical context is probably automatic, and little, if any, search process is needed. If generation produces better memory than reading because of this initial effortful search process, then we would expect to see a generation effect only under conditions in which subjects know they are generating items that are represented in the mental lexicon; otherwise, no lexical search process is likely to be initiated.

To test this account of the generation effect, half of the subjects in Experiment 1 were told that the experimenter-produced "nonwords" were, in fact, low-frequency words of the type that expert Scrabble players might use. Each of these subjects was told that he or she might or might not know the meaning of a given word. A second group of subjects was told that they would be receiving only nonwords. If knowledge about the lexical status of the items is important, then we should see a generation effect when subjects believe that they are generating actual words even though the words, in fact, are experimenter created. In short, these subjects may be induced to search the lexicon during the generation process to find a "word" that meets the generation requirements. In contrast, if representation in the lexicon is a necessary condition for the generation effect, then no generation effect should be found for either group.

Method

Subjects and Apparatus. The subjects were 32 undergraduates enrolled in an introductory psychology class at the University of Texas. Participation counted toward partial fulfillment of course requirements. All stimuli were presented via a Kodak Ektagraphic slide projector. The subjects were seated in a darkened room, in groups of one to four, facing a rear-projection screen. A Lafayette four-bank timer was used to control stimulus presentation, and a digital watch was used to time the distractor and recall tasks. The subjects responded in individual booklets provided by the experimenter.

Stimulus materials. The stimuli were four-letter nonword items arranged in rhyming pairs (e.g., FEMP ZEMP; LORP JORP). Thirty pairs, each having a unique rhyming syllable, were constructed. The nonwords were created using a rhyming dictionary and finding 30 three-letter stimuli (e.g., "EMP") that had very few rhymes. Next, two different initial consonants (e.g., "P" and "Z") were assigned to each syllable to form two rhyming items (e.g., "FEMP" "ZEMP") that were neither words nor homonyms of words. This selection method was designed to create stimuli that were as meaning-free as possible.

For the read condition, the nonword items of each pair were printed side by side in their entirety ("FEMP ZEMP"). For the generate condition, the first nonword was printed in its entirety, and beside it was printed the initial letter of the rhyming nonword. The initial letter was underscored, as were the three blank spaces following it, to indicate that the word should be completed ("FEMP Z__"). In all cases, the pairs were centered on a slide, with uppercase lettering. Additionally, slides with three-digit numbers non-divisible by three were used for a backward-subtraction distractor task. A blue cellophane slide was used as a signal to turn to a blank page in the response booklet; a blank slide was projected during the recall period.

Design. A $2 \times 2 \times 5$ (group $\times$ input condition $\times$ trial) mixed design was employed. The design was between subjects with respect to the information given about the items. One group was told that all items were experimenter-created nonwords; the other group was led to believe that all items were low-frequency English words. In-
put condition (read vs. generate) and trial were within-subject variables. The materials were counterbalanced to ensure that each word pair occurred in both the read and generate conditions for each group.

Procedure. One group of subjects (N = 16) was informed that the experiment was designed to test the way people learned experimenter-created nonwords whose visual and pronounceable characteristics mimicked English. The other group (N = 16) was told that the experiment tested how people learned words that occurred rarely in everyday experience. Specifically, the items were introduced to these subjects as words of the "type that expert Scrabble players might use to get a lot of points and win a game." These subjects were told that they might recognize some of the words, but not to worry if they did not.

Prior to the beginning of the experimental trials, the experimenter showed examples of the read and generate pairs while reading the instructions aloud. A short practice session followed. The experimenter explained that some pairs could simply be read since both the stimulus and response terms were spelled entirely (read condition), but that the response term would have to be produced for some pairs (generate condition). The subjects were told to produce the response term by using the underscored letter on the slide to form the rhyme of the stimulus term, and to spell that rhyme like the stimulus term. The subjects were instructed to read the stimulus term of the pair and then to write the read or generated response term in their booklets. The list of 30 nonword pairs was presented at a rate of 5 sec per pair. Read and generate pairs were ordered randomly in the list sequence.

Immediately following the final pair, a 5-sec blue slide signaled the subjects to turn to a blank page in their response booklets. Next, a three-digit number was displayed from which the subjects were to subtract by threes and to write down each successive response. After 30 sec of this distractor task, a blank slide was projected and the experimenter said "Recall." This was the subject’s cue to write on a blank page the response terms from as many of the pairs as possible. The experimenter encouraged guessing and stressed the unimportance of recall order. The recall interval lasted 3 min. This entire study-test sequence was then repeated four additional times with the same items rearranged randomly prior to each study sequence.

Results and Discussion

Across the five study-test trials, the subjects were quite proficient at generating the appropriate response terms. On the first study-test trial, for example, the subjects wrote correct responses for 98% of the generate pairs and for 99% of the read pairs (i.e., for which the subject simply had to copy the response term into the test booklet). For the entire session, performance was above 99% for both pair types. If either a generation or a copying error was made, the subject’s actual written response was used as the ‘‘correct’’ answer to score the free-recall test.

The results of principal interest are shown in Figure 1. Figure 1 shows recall performance for each of the five study-test trials, separated according to pair type (read vs. generate) and subject group (‘‘words’’ vs. ‘‘nonwords’’). An overall analysis of variance (ANOVA) on the data revealed only a significant main effect of trial [F(4,120) = 84.52, MSE = 3.51, p < .01] and a significant interaction between the trials variable and the read/generate variable [F(4,120) = 4.58, MSE = 1.11, p < .01]. None of the other main effects or interactions approached significance.

The left panel of Figure 1 displays the data pattern for the subjects who were aware that the to-be-remembered items were artificial, experimenter-created nonwords. The subjects in this group were not expected to show a generation effect, since their procedure followed, in most respects, the one employed by McElroy and Slamecka (1982). Collapsed across the five study-test trials, this expectation was confirmed: The subjects recalled a mean of 3.9 read items, compared with 4.2 generate items [t(15) < 1]. The right panel of Figure 1, which shows the data for the subjects who were led to believe that the items were real words, presents a similar picture: Overall, for the five study-test trials, these subjects recalled an average of 3.8 items in both the read and generate conditions. Clearly, no generation effect was obtained for either group. Although the read/generate variable did interact with trials, the nature of that interaction is neither clear-cut nor consistent with any straightforward accounts of the generation effect. Consequently, these results suggest that knowledge about the lexical status of the to-be-remembered items was not an important determinant of McElroy and Slamecka’s failure to find a generation effect with nonwords. The subjects in the ‘‘word’’ group were led to believe that the stimulus and response terms were actual English words, yet no generation advantage was produced with these subjects. Thus, if these subjects were induced to attempt a search of the lexicon during generation, then that search had no apparent effect upon retention. Of course, it is possible that the task instructions were not sufficient to convince these subjects completely that the nonwords were, in fact, low-frequency words. Yet, it seems reasonable to argue that the ‘‘word’’ subjects would have made some attempt to check for lexical entries, at least during the first study-test trial.

In short, this experiment verifies that the act of generation, by itself, is not sufficient to produce a retention advantage over read stimuli. The data are therefore con-
sistent with interpretations that implicate semantic or lexical memory in the generation process. Equally so, the data are inconsistent with viewpoints that appeal to the "effortful" nature of generation: As McElroy and Slamecka (1982) argued, the effort hypothesis, in its simple form, predicts that the generation advantage should not depend importantly upon the lexical status of the stimuli. Production of a response should be inherently more "effortful" than reading regardless of the items employed and, consequently, should yield better memory. In contrast to this position, the current data suggest that the generated item must be represented in the mental lexicon in order for the generation process to be reflected in retention.

**EXPERIMENT 2**

One important characteristic of items that are represented lexically is that they are, in some sense, meaningful to the subject. It is possible to process a word semantically, but this is less true, if at all, for a nonword. Our purpose in Experiment 2 was to ask whether simple assignment of "meaning" to a nonword would be sufficient to make it more memorable when generated rather than read. One could account for such a finding by assuming that generation, regardless of the particular rule employed, automatically activates an item's semantic properties (if present), which then serve to aid subsequent retention (e.g., Craik & Lockhart, 1972). Stated differently, when a subject generates a "meaningful" item from "meaningless" cue fragments, he or she may simply be induced to think about the item semantically. For nonwords, of course, since the items possess no meaning, this kind of obligatory semantic processing is not possible. Such reasoning suggests that it might be possible to produce a "generation effect" when nonwords acquire meaning episodically (that is, become "meaningful" items to the subjects).

To investigate this idea, the subjects in Experiment 2 either generated or read experimenter-created items that had first been assigned semantic properties. As in Experiment 1, the generation rule was rhyme based. However, in Experiment 2, half of the stimulus-response pairs (e.g., JORP LORP) contained items with previously learned definitions; the remaining pairs were left undefined. Once again, if generation, regardless of the particular production rule employed, tends to induce a kind of obligatory semantic processing, then one might expect to see a "generation effect" for the defined pairs because subjects are given the task of generating "meaningful" response items (albeit nonwords) from "meaningless" cue fragments.

**Stimulus materials.** A set of 12 nonword pairs was selected from those employed in Experiment 1. In order to reduce interitem similarity even further, some pair members were reversed, and the initial letter of other items was changed. Each of the 12 nonword pairs was prepared in both a "read" and a "generate" form on projection slides. For the first stage of the experiment, in which subjects learned definitions for some of the nonwords, the subjects were exposed to study lists of all 24 nonwords that constituted the 12 pairs, randomly arranged. Both the stimulus and response members for 6 of the pairs were assigned definitions; the 12 items from the remaining 6 pairs were given no definitions, but were followed by a row of asterisks. Thus, subjects were exposed to all 24 items during the initial learning phase. The definitions were category names arbitrarily selected from the Batitt and Monague (1969) and Shapiro and Palermo (1970) category norms. Four different random orders for the study and test lists were prepared to ensure that the subjects would not learn the definitions by rote memorization of list location. In addition, the set of 6 pairs given definitions was counterbalanced across subjects.

**Design.** A $2 \times 2 \times 2$ (input condition x meaning x trial) mixed design was employed. The input condition (read vs. generate) variable was between subjects; meaning (defined vs. undefined) and trial were both within-subject variables. Overall, the nonword pairs were counterbalanced such that each nonword pair appeared in both the read and generate form, and, within each of these two types of tasks, each pair was defined and undefined equally often across subjects.

**Procedure.** The subjects were tested individually due to the nature of the learning task. Each subject was told that the items presented would be "nonwords," but that an attempt was being made to create real words by assigning definitions to some of the items. A study-test procedure was employed in which 3 min were allowed for study, followed by a self-paced test. For the test, the subjects were instructed to write the assigned meaning beside each defined nonword and an asterisk beside those items that had been left undefined. Twelve of the 24 items were defined. The subjects were told that they might need to study the list as many as three or four times before their recall reached the criterion of two perfect trials in succession. Guessing was encouraged, as was a quick test pace. In order to facilitate learning, the subjects scored their own tests while the experimenter watched. The study interval following the first perfect recall was reduced to 1.5 min. The four forms of both the study and test list were rotated such that a different form was given on each successive trial until the subject had reached criterion.

After they had reach criterion, the subjects were told that the items from the learning task would appear in pairs on projection slides. The generate group (N = 18) received rhyme-generation instructions as in Experiment 1. The read group (N = 18) was told simply to read the pair and copy down the second member. Both groups received a short practice session with nonwords not previously seen. Pairs were presented individually, at a rate of 4 sec per pair. For both groups, the 30-sec subtraction task and a 1-min recall task followed immediately after the 12th pair. For the recall task, as in the previous experiment, the subjects were told to write down the second member from as many pairs as possible; guessing was encouraged. All subjects received the same 12 pairs, randomly rearranged prior to the second trial. A final test for the assigned definitions was administered after the second trial.

**Results and Discussion**

In the first stage of the experimental session, in which the subjects learned definitions for half of the nonword pairs, the criterion was set at 2 perfect trials in succession. For the 36 subjects combined, a mean of 4.1 study-test trials was required to meet this criterion. In the sec-
ond stage of the session, in which the subjects either copied or generated and wrote the response terms, performance was nearly perfect. Across the 2 generate/read trials, only three errors were made (one generation error and two copy errors).

The left panel of Figure 2 shows the critical free-recall data as a function of meaning and input condition. Of initial interest is the main effect of meaning, in which the defined nonwords were recalled significantly better than the undefined nonwords [F(1, 34) = 21.46, MSe = 2.28, p < .01]. This finding presumably reflects the greater amount of processing that the defined nonwords received during the first stage of the experiment (although the nominal amount of exposure was the same for both defined and undefined items). Second, the main effect of input condition was not significant; that is, there was no evidence for a generation effect [F(1, 34) < 1]. Most importantly, there was no significant interaction between input condition and meaning [F(1, 34) < 1]. The only significant effects were trials [F(1, 34) = 45.99, MSe = .44, p < .01] and the interaction of trials with input condition [F(1, 34) = 6.31, MSe = .44, p < .05]. Analysis of this interaction, however, provides no supportive evidence for the presence of a generation effect. Input condition appeared to have some effect on retention for the first study-test trial (for the defined nonwords), but the pattern is in the direction of an advantage for the read, not the generate, pairs.

Consequently, the results indicate that “artificially” providing a nonword with semantic attributes is not sufficient to produce a pattern of results that is typical of the generation effect. Despite this conclusion, it is interesting to note that the subjects often reported after the session that they had thought about the meanings of the defined words during the read/generate portions of the experiment. In other words, although no systematic observations were made, casual discussions suggested that the subjects had treated the defined items as “meaningful” throughout the later stages of the session. This observation is supported by the results of the definition test at the end of the experiment. Collapsed across the read and generate groups, the subjects produced the correct definitions for 96% of the nonwords. These results verify the success of the initial learning phase and show that the meaningful attributes of the stimuli were at least available to the subjects throughout the read/generate portions of the experiment.

As a further test of the hypothesis under study, Experiment 2 was replicated using a completely within-subject design (N = 24). The study was divided into three phases: In Phase 1, all subjects learned to associate definitions with a set of 12 nonwords. In Phase 2, the subjects were required to generate, from rhyming nonword cues, 6 defined nonwords and 6 undefined nonwords (no stimulus terms received definitions in this experiment); they also “read” 6 defined nonwords and 6 undefined nonwords. In Phase 3, the subjects were asked to free-recall all of the generated and read items. The results from the final free-recall test are shown in the right panel of Figure 2. As in Experiment 2, the defined items were remembered significantly better than the undefined items, but no evidence for a generation effect was found for either item type; in addition, the slight advantage for the defined “read” items noted on the first test trial of Experiment 2 was no longer apparent. Specifically, for the undefined items, 10 subjects recalled more generated nonwords than read nonwords, 9 subjects showed the reverse pattern, and there were five ties. For the defined items, 8 subjects recalled more generated than read nonwords, 6 did the reverse, and there were 10 ties. Clearly, these data provide no support for a generation effect.

**EXPERIMENT 3**

The results of Experiment 2 indicate that it is not the mere possession of semantic properties that determines whether generation of an item will benefit performance. Perhaps more importantly, if one assumes that the defined nonwords were, in some sense, represented in the lexicon as a consequence of the initial training (see Shiffrin, Salasoo, & Feustel, 1984), then the data of Experiment 2 also suggest that lexical representation itself may not be a sufficient condition to produce the generation advantage. Instead, in Experiment 3, we investigated the idea that the critical factor underlying the generation effect may simply be how related the to-be-generated item is to other information in the memory system: Words, by virtue of their established semantic properties, are associatively related to other lexical items, whereas nonwords possess few, if any, associative relations. Thus, although the initial training of Experiment 2 may have been sufficient to establish weak lexical entries for the defined nonwords, it is unlikely that those items were part of any established retrieval networks. The following account of the generation effect is based upon two assumptions: First, generation results in greater activation of an item’s location in
the lexical network than does reading. Second, this activation spreads, where possible, to associated entries, thereby creating a network of potential retrieval routes for the activated item. Since generation produces more activation of an item’s lexical representation, a greater retrieval network should be established when an item is generated. For nonwords, even though more cognitive processing may still be required during generation, the activation is likely to be restricted to the item itself; that is, there are few, if any, related items to receive added benefit from the enhanced activation, so no generation advantage is found.

These ideas suggest that demonstrations of the generation effect should not hinge exclusively upon lexical representation, but rather upon variables that tap the level of association that the generated item has to other items in the network. One such variable is word frequency—high-frequency words are thought to have greater “spread” (more associates) than low-frequency words (cf. Postman & Keppel, 1970). If representation in the lexicon is a necessary and sufficient condition to produce a generation effect, then one would expect to see a generation effect for words, regardless of frequency, and no generation effect for nonwords. If, however, the effect depends upon an item’s number of associates, then we might expect the read/generate variable to interact significantly with frequency of use in the language. This hypothesis was investigated in Experiment 3: Subjects were asked to read and generate nonwords and words with three levels of word frequency (low, medium, and high). The generation task differed from that in the previous two experiments in that the subjects were asked simply to switch the first two letters of an item, if underlined, and to copy the item intact if no letters were underlined. For example, the stimulus “EHAVEN” would, when generated, become “HEAVEN,” whereas the stimulus “SPIRIT” would merely be copied into the response booklet. Retention for the read and generated items was assessed at the end of the experiment with a yes/no recognition test.

**Method**

**Subjects and Apparatus.** Thirty undergraduate subjects from the University of Texas participated for course credit. The subjects were tested in groups of three or fewer with apparatus similar to that described in Experiment 1.

**Materials and Design.** All subjects received the same 48 to-be-remembered items, which were composed of 12 nonwords, 12 low-frequency words (Thordike & Lorge, 1944, rating of 1 or less per million), 12 medium-frequency words (25–40 per million), and 12 high-frequency words (AA, or greater than 100 occurrences per million). All four groups of stimuli were matched for letter length, and the words were matched for mean imagery value (mean “T” value was 3.99; cf. Palvio, Vuille, & Madigan, 1968). The four sets of stimulus items are shown in Table 1.

A 4 × 2 (item type × input condition) within-subject design was employed. Each subject generated, through switching the first two letters, 24 items (six instances of the four item types) and read 24 items (six instances of the four item types). Generate items were presented with the first two letters underlined (gaven); read items were presented intact and with no letters underlined (heaven). Across subjects, each item appeared equally often in read and generate forms. On the final recognition test, the subjects attempted to discriminate the 48 study items from 48 distractors of the same item types: that is, 12 nonwords and 12 low-frequency, 12 medium-frequency, and 12 high-frequency words. On the recognition test, each study item was presented intact (as the product of generation or, if read as originally presented) and with no letters underlined. All subjects received the same final test.

**Procedure.** The subjects were asked to copy or generate each item dependent upon whether the first two letters of the presented item were underlined. Thus, although the generation procedure formed a meaningful item for only the words, it could be applied equally well to the nonwords; the subjects merely had to switch the underlined letters. Each item was presented via a slide projector onto a blank wall for 6 sec. During this period, the subjects were expected to switch the underlined letters and write the resulting item in a response booklet. If no letters were underlined, the subjects simply copied the item into the test booklet. For both item types, the subjects were asked to try to remember what they had written. Read and generate items were intermixed randomly throughout the study sequence. A short practice session preceded the study session.

Following the last study item, a three-digit number appeared, from which the subjects were asked to count backward by threes. The subjects wrote their counting responses on a blank page for 1 min. After the distractor task, the subjects were handed a three-page recognition test booklet. Each page contained 32 test items composed of 16 “old” items and 16 distractors; each condition of the experiment was represented equally on each page of the test. The subjects were told simply to circle those items that they recognized from the study phase of the experiment.

**Results and Discussion**

Performance on the letter-switching generation task and the read (copy) task was virtually perfect. Across the 30 subjects, only seven errors were made on the generation task (of 720 generation opportunities), and only three errors were made on the copy task. The results of main interest are shown in Figure 3. Figure 3 shows mean percent correct recognition performance for the read and generate conditions as a function of item type. The number of correct recognitions served as the index of recognition because the different item types were intermixed randomly within the recognition test and thus there was little reason to expect differences among the read and generate conditions to reflect response bias. The highest false-alarm rate was found for the nonwords (.13); there were no differences in the false-alarm rates for the low- (.04), medium- (.04), and high-frequency (.05) words.
An ANOVA on the data summarized in Figure 3 revealed a significant main effect of input condition \( F(1,29) = 11.72, \text{MSe} = 1.32, p < .01 \), a significant effect of item type \( F(3,87) = 4.42, \text{MSe} = 1.49, p < .01 \), and a significant interaction \( F(3,87) = 11.93, \text{MSe} = 0.74, p < .01 \). As in the previous two experiments, no advantage was found for generating instead of reading nonwords. For this condition, which is displayed on the far left side of Figure 3 ("N"), 8 of the subjects correctly recognized more of the generated nonwords, 13 subjects recognized more of the read nonwords, and there were nine ties. Once again, no generation effect was found with experimenter-created items.

Of far more interest in the present context are the data from the word stimuli that differed in frequency level. First, examination of the data from the read condition reveals the standard word frequency effect: The low-frequency words ("L") were recognized better than the medium-frequency words ("M"), which were recognized better than the high-frequency words ("H"). Although this pattern is not surprising, it verifies that we functionally manipulated frequency for the subjects in the experiment. Second, a significant generation effect was found for the medium- and high-frequency words but not for the low-frequency words. Specifically, for the low-frequency words, 7 subjects recognized more generated items, 5 subjects recognized more read items, and there were 18 ties (only 6 of the ties were at ceiling). Clearly, these data provide no hint of a generation effect for these stimuli. For the medium-frequency words, 18 subjects recognized more generated items than read items, 3 subjects recognized more read items, and there were 9 ties \( p < .001 \). For the high-frequency words, 21 subjects recognized more generated items correctly, 3 subjects recognized more read items, and there were 6 ties \( p < .001 \).

The fact that low-frequency words did not produce a generation effect in this experiment is important theoretically because it indicates that representation in the mental lexicon is not a sufficient condition for producing a generation advantage. It seems reasonable to assume that the low-frequency words were represented in the subjects' vocabularies, yet generation did not enhance retention. In support of this conclusion, a separate ANOVA comparing the nonwords and the low-frequency words revealed a significant difference between the two item types: The low-frequency words were recognized significantly better than the nonwords \( F(1,29) = 5.69, \text{MSe} = 1.50, p < .03 \), whereas neither input condition nor the interaction was significant. This finding reduces the likelihood that the low-frequency words were functionally identical to the nonwords in this experiment. Rather, it appears that low-frequency words, like nonwords, simply do not produce generation effects. As such, the results are consistent with the main hypothesis of Experiment 3: The critical factor underlying the generation effect may be the network of associations that an item has with other represented items in the lexicon. If an item is represented lexically, but is linked to relatively few other meanings or associates, then its generation may not be that beneficial to the subject. Nonwords are simply a special case of items with few or no lexical associations.

It is also of interest to note that no differences were found in the size of the generation effect between the medium- and high-frequency words. A separate ANOVA on these item types revealed significant effects of input condition \( F(1,29) = 35.32, \text{MSe} = 1.12, p < .001 \), and item type \( F(1,29) = 6.25, \text{MSe} = 0.97, p < .03 \), but the interaction was not significant \( F(1,29) < 1 \). Thus, although the generation effect clearly is influenced by frequency, the exact nature of that relationship cannot be determined from the present results. It is possible that an item must possess some minimal amount of "spread" (or number of associates), as indexed by frequency, for generation to be beneficial, but such reasoning is speculative. In addition, since frequency is typically confounded with a number of other item characteristics (see Gerbacher, 1984), a more extensive parametric analysis needs to be undertaken—for example, it is necessary to disentangle frequency from variables like familiarity in future experiments. For our immediate purpose, however, these results remain important because they disconfirm the lexical representation hypothesis of McElroy and Sliamacka (1982).

**GENERAL DISCUSSION**

The preceding experiments were designed to investigate the finding of McElroy and Sliamacka (1982) that the "generation effect" depends importantly upon the lexical status of the to-be-remembered items. In three separate experiments, the generation of items not present in the subjects' existing vocabularies failed to produce a retention advantage over a simple copying condition. Moreover, this conclusion held across wide variations in the overall retention levels of the to-be-remembered items.
For example, the "defined" nonwords in Experiment 2 were recalled at significantly higher levels than the undefined items, yet no generation effect was produced in either case. Likewise, similar performance for the generate and read conditions was found across repeated study-test trials, for which the overall retention levels increased dramatically. Finally, in Experiment 3, comparable retention levels were obtained for nonwords and medium- to high-frequency words, yet only the latter items produced significant generation effects. The last result also indicates that appeals to the specific generation rules used in these experiments (i.e., rhyming or letter switching) cannot explain the null effect for nonwords; exactly the same letter-switching rule was used for all items in Experiment 3, yet only medium- and high-frequency words showed effects of the generation task. From a data standpoint, therefore, these experiments provide a strong replication of the basic pattern reported by McElroy and Slamecka (1982).

More importantly, the present experiments were intended to fine-tune the boundary conditions surrounding the failure of nonwords to produce the generation effect. In Experiment 1, some subjects were led to believe that the nonwords were, in fact, words. We reasoned that these subjects might be induced to search the lexicon in an attempt to find some meaning for a given item. Such a process may not have occurred in the work of McElroy and Slamecka (1982), since their subjects were aware of the artificial nature of the stimuli. Despite the plausible benefits of an attempted search (e.g., greater processing effort, greater likelihood of elaborations with other items in memory, etc.), knowledge about the lexical status of the items had no impact upon the generation effect. Experiment 2 sought to create meaningful stimuli episodically in order to see whether the assignment of semantic attributes alone would be sufficient to produce a generation advantage (i.e., to make nonword items act like words). It is possible, for example, that any kind of generation task acts automatically to activate semantic features because subjects typically are asked to generate a meaningful item from a meaningless cue fragment. Despite assurances that the subjects were able to retrieve the assigned definitions, a generation effect was still not produced. Finally, Experiment 3 showed that, although representation in the mental lexicon may be a necessary condition for production of the generation effect, as argued by McElroy and Slamecka, it is not a sufficient condition. In addition to nonwords, low-frequency words failed to receive any benefit from generation for later retention.

To account for these findings, we would like to offer a very simple description of the mechanisms that may lie behind the generation advantage. One need assume only that generation, perhaps because it is more "effortful," activates an item's location in the lexical network to a greater extent than does reading. This activation then spreads to related entries, enhancing the activated item's episodic retrieval network; the end result is that a greater number of functional retrieval routes are established for generated items. This difference in activation level between generating and reading is felt to hold across a wide variety of generation rules (rhyme, orthographic, semantic) and stimulus types (nonwords, words). Greater activation is a by-product of generation, regardless of the particular production rule employed. If an item is not represented in the lexicon, then, of course, no generation effect will be found. If an item is represented, but has few associated entries, then the effect of generation is less likely to be reflected in performance because there are few potential retrieval routes capable of being established from any enhanced activation. Thus, low-frequency words, although clearly represented lexically, do not produce a generation effect.

In summary, the three experiments reported here replicate and extend the finding of McElroy and Slamecka (1982) that the generation effect is not obtained when nonwords are used as the to-be-remembered items. Although each experiment helped to eliminate alternative explanations of the finding, and thereby increased its generality, a number of questions remain unanswered. It would be of special interest, for example, to determine whether the failure to obtain a generation effect with nonwords and low-frequency words is due to the special properties of the stimuli themselves (i.e., low levels of familiarity?) or whether it is best to describe these stimuli as endpoints of a continuum expressing how related an item is to other potential retrieval cues in the memory system.

REFERENCES


(Manuscript received December 11, 1984; revision accepted for publication February 18, 1985.)