Journal of Educational Psychology
1987, Vol. 79, No. 2, 162-170

Copyright 1987 by the American Psychological Association, Inc.
0022-0663/87/$00.75

Effects of Variable Encoding and Spaced Presentations on Vocabulary Learning

Frank N. Dempster
University of Nevada, Las Vegas

I examined the applicability of the encoding variability hypothesis and the spacing phenomenon to vocabulary learning in five experiments. I manipulated encoding variability by varying the number of potential retrieval routes to the word meanings, using a one-sentence context condition, a three-sentence context condition, and a no-context (definitions-only) control condition. I evaluated the spacing effect by presenting each word with or without intervening words. The results provided no evidence that the opportunity to establish multiple retrieval routes by means of contextual information is helpful to vocabulary learning, a conclusion supported unequivocally by all five experiments. By contrast, spaced presentations yielded substantially higher levels of learning than did massed presentations. I discuss the results largely in terms of educational concerns, including the utility of the learning-from-context approach to vocabulary learning.

In the experiments reported in this article, I investigated the applicability of two principles derived from traditional verbal learning research—variable encoding and the spacing effect—to a real-world activity, vocabulary learning. More specifically, the principal questions under investigation were whether vocabulary learning would be enhanced by (a) the opportunity to establish multiple retrieval routes to word meanings and (b) spaced rather than massed presentations of the to-be-learned information.

Background and Empirical Evidence

Two of the best known concepts to emerge from verbal learning research are encoding variability, which refers to the proposition that the probability of recall varies directly with the number of retrieval routes, and the spacing, or lag, effect, which refers to the advantage of spaced over massed presentations. In the late 1960s and early 1970s, these concepts were the source of growing fascination, as evidenced by a rapidly growing body of literature emphasizing their centrality and importance to theories of learning (Bower, 1972; Hintzman, 1974; Martin, 1972; Melton, 1970). For example, Martin listed a variety of phenomena that he argued, seemed inexplicable except in terms of encoding variability, and Hintzman stressed the notion that a phenomenon as ubiquitous as the spacing effect must be telling something important about memory. In general, the findings associated with multiple encodings and spaced presentations seemed to constitute an important breakthrough in the quest to understand the foundations of effective learning.

As might be expected, this attention soon attracted the interest of those looking for research products with application potential (e.g., Bjork, 1979; Clifford, 1981, p. 326). Foremost among these is Bjork, who recommended the application of the encoding variability hypothesis and the spacing effect to various facets of instruction, including study methods, course design, and lecture design. In the area of course design, for example, Bjork argued that there was enough evidence to suggest that students would be better served if they were given the opportunity to encode an instructional objective in multiple contexts instead of in a fixed context and if the presentation format was spaced rather than massed.

In addition to informing such outright prescriptions, the encoding variability hypothesis appears to be implicated in certain already established teaching methods, most notably the learning-from-context approach to vocabulary learning (e.g., Johnson & Pearson, 1984; Sternberg, Powell, & Kaye, 1983). Although there are several variations of this approach, it generally consists of having students attempt to infer the meaning of a new word from the contextual cues provided by sentences in which the word is used. Moreover, it is recommended that the sentences in which the word is used constitute a variety of contexts so that the concept represented by the word can more easily be related to a variety of other concepts. In other words, several sentences are better than one. By contrast, Sternberg et al. (1983) suggested, "Mere repetition of an unknown word in essentially identical contexts is unlikely to be helpful, because few or no really new cues are provided regarding the word's meaning" (p. 224).

A careful reading of the literature, however, suggests two rather serious reservations about these recommendations. One reservation stems from the fact that whereas early studies provided strong empirical support for both principles, subsequent work has called into question their reliability. Regarding encoding variability, most of the more recent studies either

Preparation of this article was supported in part by a grant from the Barrick Research Council, University of Nevada, Las Vegas. Portions were presented at the 1986 annual meeting of the American Educational Research Association.

I thank Deborah K. Boland for her assistance in collecting the data and my colleagues John A. Bates, Mark Beals, and Philip Steedman and two anonymous reviewers for their helpful suggestions for improving an earlier draft of the article.

Correspondence concerning this article should be addressed to Frank N. Dempster, University of Nevada, College of Education, Las Vegas, Nevada 89154.

have failed to support the theory or have yielded extremely weak effects (see Postman & Knecht, 1983, for a thorough review). The situation with respect to the spacing of repetitions, a phrase that Melton (1970) made famous, is not quite as bleak. The spacing effect is still one of the most reliable phenomena in experimental psychology. Recent studies using standard verbal learning materials have shown, however, that it is subject to certain boundary conditions that are not fully understood (e.g., Glenberg, 1977; Toppino & Gracen, 1985).

The second reservation concerning the applicability of encoding variability and the spacing effect to school-related activities is that in most cases, the tasks (e.g., memorizing lists of unrelated words or sentences) in which they have been found effective are not representative of school-related activities. With respect to the spacing of repetitions, most studies using more naturalistic activities (typically, text-processing activities) have either failed to reveal lag effects (Ausubel, 1966; Gay, 1973, Experiment 1; Peterson, Ellis, Toohill, & Kloess, 1935; Sones & Stroud, 1940), or yielded rather weak effects (English, Wellborn, & Killian, 1934; Gay, 1973, Experiment 2). Studies reporting robust effects of spacing on tasks that have clear classroom analogues are the exception rather than the rule (Dempster, 1986; Reynolds & Glaser, 1964). Overall, these studies hardly inspire much confidence in the practical utility of the spacing effect.

Regarding encoding variability, the few educationally relevant studies are the ones testing the merits of the learning-from-context approach to vocabulary learning. In these studies, a condition in which new words accompanied by sentences in which the words were used was compared with some form of a definition (i.e., no-context control) condition (Christ & Petrone, 1977; Gipe, 1979; McDaniel & Pressley, 1984). In two of these, subjects who were presented with words in several different sentence contexts performed better than subjects who were simply presented with definitions (Christ & Petrone, 1977) or who looked up the new word in a dictionary (Gipe, 1979). By contrast, in another study, subjects presented with definitions performed considerably better than did subjects who were presented with words in the context of three sentences (McDaniel & Pressley, 1984).

One perhaps key difference among these studies is that in the first two, the sentence contexts for each word were apparently quite different from each other, whereas in the most recent study, each sentence for a particular word provided the same or a similar context. Thus, it is not clear that McDaniel and Pressley's (1984) manipulation would be expected to provide multiple retrieval routes to the to-be-learned words. From the present perspective, another factor complicating the interpretation of McDaniel and Pressley's results is that subjects in the context condition were not provided with definitions during learning even though the dependent measure was the number of meanings recalled. Finally, neither McDaniel and Pressley nor any other researcher who compared definitions with sentences included a comparison between a single-sentence context condition and a multiple-sentence context condition. Given the assumption of advocates of the learning-from-context approach to vocabulary learning that several sentences are better than one (e.g., Sternberg et al., 1983), such a comparison would be meaningful.

In sum, the evidence reviewed in this section casts serious doubts on the proposition that the principles of encoding variability and the spacing effect can be used to promote more effective learning in real-world situations. These doubts stem from questions that recent studies have raised about the reliability of these principles in traditional verbal learning tasks and about the need for further studies using school-related tasks. I designed the present experiments to expand understanding of the applicability of encoding variability and the spacing effect to one such task.

Present Experiments

In the present experiments, I examined the effects of variable encoding and spaced versus massed presentations on the acquisition of unfamiliar words and their meanings. The principal dependent measure was the number of word meanings recalled.[1] I manipulated encoding by varying the number of potential retrieval routes to the word meanings. In the no-context control condition, the words were presented with their definitions only (Experiments 1–5). In the single-sentence context condition, each word was also used in one sentence (Experiments 1 and 2), whereas in the three-sentence context condition, each word was also accompanied by three sentences in which it was used (Experiments 1–5). In Experiment 1, each word was presented once, as is often the case in vocabulary learning research, whereas in Experiments 2 through 5, each word was presented three times. I evaluated the spacing effect in Experiments 3 and 4. In these experiments, each of three presentations of a word was either massed, with each repetition occurring in succession, or spaced, with each repetition separated by the presentation of other words. In Experiment 5, each repetition was spaced, and a no-presentation baseline condition was included. The design of each experiment is summarized in Table 1.

Experiment 1

Method

Subjects. Subjects were 36 students enrolled in undergraduate educational psychology courses at a state university and received extra course credit for their participation. They were randomly assigned to conditions, resulting in 12 subjects per condition.

Materials. Thirty-eight uncommon English words were selected as the to-be-learned vocabulary. Eight of these words were used by McDaniel and Pressley (1984), and the remainder were selected by several graduate students and me. We selected a word primarily if we judged it unlikely to be known by the sample. The list consisted of 26 nouns, 9 adjectives, 1 adverb, 1 verb, and 1 preposition.

[1] The ability to write sentences in which words are used properly might be assumed to be a preferable dependent variable. McDaniel and Pressley (1984), however, found that this ability was determined by whether subjects knew the meaning of the words (i.e., could write definitions of the words) in the first place, even among subjects in their context condition. Hence, the present dependent variable seems a perfectly adequate measure of vocabulary learning.

Table 1
Summary of the Design of Experiments

Condition	Experiment 1	2	3	4	5
Spacing					
Massed	NA		X	X	
Spaced	NA	X	X	X	X
Encoding					
Control (definitions only)	X	X	X	X	X
Definition + one sentence	X	X			
Definition + three sentences	X	X	X	X	X
No-presentation baseline					X

Note. NA = not applicable; X = applicable.

Nine 38-page booklets, with one vocabulary word per page, were constructed. There were three booklets per condition. The booklets for the no-context control condition contained only the words and a one- to four-word definition of each. The booklets for the single-sentence context condition, in addition to providing the definitions, provided one sentence in which each word was used. The booklets for the three-sentence context condition provided the definitions plus three different sentences in which each word was used. An equal number of subjects (4) in the single-sentence context condition received booklets containing each sentence used in the three-sentence context condition. In constructing the sentences for any given word, every effort was made to provide different semantic contexts for the word so that the three-sentence condition would truly provide for variable encoding. Table 2 provides an example of a booklet entry for each encoding condition. A cassette tape was used as a companion to the booklets. The recording signaled the subjects when to open their booklets and when to turn a page.

Procedure. Subjects were tested either individually or in groups ranging from 2 to 6 in a room ordinarily used for counseling. All subjects were told that they would be presented with 38 vocabulary words and that most, if not all, words would be unfamiliar to them. In addition, it was explained that each word, typed in uppercase letters, and its definition would appear on a separate page in booklets that they would receive and that the definitions were typed immediately to the right of the words. It was emphasized that their task was to attempt to learn the meaning of each and that if there was any other information on the pages, they were to use it in trying to learn the words' meanings. Finally, subjects were informed that they would have 20 s to study each word and that a prerecorded tape would pace them accordingly. The command "Begin" was their signal to open their booklets, and the command "Turn," occurring at 20-s intervals, was their signal to turn a page.

When the tape was completed, the booklets were returned to folders, and the subjects were instructed to count backwards by threes aloud for 1 min from a three-digit number read by the experimenter. Following this distractor activity (intended to minimize recall from short-term memory), a recall test was administered. For the recall test, each subject was given a 2-page test booklet, which they were instructed to open immediately following the distractor activity. The booklets contained all 38 vocabulary words (in a random order different from the one in which they were initially presented) in a column followed by blanks for the definitions. Subjects were allowed 10 min for recall.

Table 2
Examples of Encoding Conditions

Encoding condition	Presentation format
No-context control	LOGGIA—balcony
One-sentence context	LOGGIA—balcony 1. Juliet stood on the *loggia* while Romeo declared his love.
Three-sentence context	LOGGIA—balcony 1. Juliet stood on the *loggia* while Romeo declared his love. 2. The upper *loggia* at the opera house was filled to capacity. 3. Each apartment had its *loggia* overlooking the courtyard.

Note. The vocabulary word LOGGIA was selected for illustration because McDaniel and Pressley (1984) also used it. Thus, the reader can compare the present sentences with theirs.

Results and Discussion

For purposes of scoring, either definitions as presented or definitions that paraphrased the presented definitions were acceptable. For example, the presented definition for APOCRYPHAL was "not genuine; sham." If subjects wrote either "not genuine" or "sham," they were given credit because these responses matched the presented definition. But they were also given credit if they wrote "not real," as several subjects did, because this response means the same as "not genuine." Likewise, subjects who wrote "under the table" next to SUB ROSA received credit because this definition means the same as "secretly; privately."

The mean number of definitions correct was 17.00 for no-context control, 15.33 for one-sentence context, and 16.50 for three-sentence context. Thus, there were only slight differences in recall among the three conditions. As might be expected, therefore, a one-way analysis of variance (ANOVA) revealed that with alpha set at the conventional .05 level, these differences were not significant, $F(2, 33) < 1$, $MS_e = 91.26$. In addition, a detailed examination of the recall protocols failed to reveal anything resembling reliable differences in the number of misplaced definitions (e.g., "secretly" for APOCRYPHAL) given in the three conditions. Thus, there was no evidence that the conditions of presentation significantly affected the number of definitions learned, irrespective of whether they could be placed correctly.

One might expect context-related intrusion errors to be more common after variable encoding than after fixed encoding. The larger the number of contexts in which a word appeared, the more likely it should be that context-related words should have been accessed at the time of recall. However, scoring revealed very few intrusion errors. The total number of context-related intrusions (e.g., "opera house" for LOGGIA) was zero for the no-context control, as would be expected, five for the single-sentence context condition, and two for the three-sentence condition. In fact, the vast majority of errors in this and the following experiments were errors of omission. On the basis of the subjects' actual responses, therefore, there is no reason to believe that performance in the context conditions was hindered by the retrieval of context-related information at the time of the test.

In conclusion, the results of the present experiment fail to support the hypothesis that multiple-sentence contexts lead to better vocabulary learning than does a single-sentence context or the hypothesis that the provision of a context leads to better vocabulary learning than does no context. Instead, performance in the two context conditions was less than 8 percentage points apart, whereas the context versus no-context comparisons showed a trend in the opposite direction.

Experiment 2

Many individuals, particularly those who have advocated the learning-from-context approach to vocabulary learning, might consider the results of Experiment 1 rather surprising. Thus, it was important to conduct another test of the encoding variability hypothesis. Further, it is possible that the advantage of having one or more context-provided retrieval access routes to definitions might become apparent only with repeated presentations. In most tests of the encoding variability hypothesis, subjects receive repeated presentations of the to-be-learned material. This possibility was examined in Experiment 2 by using the encoding conditions used in Experiment 1.

Method

Subjects. Subjects were 36 students enrolled in undergraduate educational psychology courses at a state university and received extra course credit for their participation. They were randomly assigned to conditions, resulting in 12 subjects per condition. None of the subjects had participated in Experiment 1.

Materials. The only changes made in Experiment 2 were in the arrangement of the booklets and the pacing rate recorded on the cassette tape. Each booklet now had 114 pages, with each vocabulary word presented three times (37 pages between each presentation). The words were presented in the order in which they were presented in Experiment 1. In the booklets for the no-context control condition, each repetition of a word was accompanied only by its definition. In the booklets for the single-sentence context control condition, each repetition of a word was accompanied by its definition and the same sentence. In the three-sentence context condition, each repetition of a word was accompanied by its definition and a different sentence. Once again, an equal number of subjects in the single-sentence context condition received each sentence used in the three-sentence context condition.

Procedure. The only procedural changes made were necessary to conform to the changes in the materials. Thus, subjects were told that their booklets contained 114 pages and that each word was repeated three times. They were also informed that they would have 7 s to study each page (in line with the change in pacing rate). This rate was chosen for equating total study time across experiments. (Subjects were estimated to need 1 s for the two additional page turns per word in Experiment 2.)

Results and Discussion

The scoring procedure was the one used in Experiment 1, and the results can be quickly summarized. The mean number of definitions correct was 19.17 for no-context control, 21.25 for one-sentence context, and 18.83 for three-sentence context. Once again, there were rather small differences in recall between the conditions, and with alpha set at .05, these differences did not approach significance, $F(2, 33) < 1$, $MS_e = 86.11$. In addition, a detailed examination of the recall protocols failed to reveal noteworthy differences in the number of misplaced definitions or any other systematic difference across conditions. Finally, the number of intrusion errors was once again too small to be of any significance.

In conclusion, the results of Experiment 2 essentially replicate those of Experiment 1 in failing to support the hypothesis that one or more sentence contexts lead to better vocabulary learning than does no context. Moreover, performance in the one-sentence condition was higher than it was in the three-sentence context condition, the reverse of what was found in Experiment 1.

What stands out most clearly in comparing the results of the two experiments, however, is that without exception, performance was noticeably higher in the various conditions in Experiment 2 than it was in Experiment 1. It is difficult to see how these differences could have been due to the slight difference in the pacing rate used in the two experiments. A more likely explanation is that the factor responsible for these differences was the repeated, spaced presentations used in Experiment 2. Whereas Experiment 1 consisted of one lengthy presentation, Experiment 2 consisted of three shorter, spaced presentations. To examine the possibility that spaced presentations per se have a facilitating effect on vocabulary learning, I compared massed presentations versus spaced presentations in the no-context control condition and the three-sentence context condition in Experiment 3. At this point, little seemed to be gained by further examination of the single-sentence context condition, so I omitted it from Experiment 3.

Experiment 3

Method

Subjects. Subjects were 48 students enrolled in undergraduate educational psychology courses at a state university and received extra course credit for their participation. They were randomly assigned to the four conditions, resulting in 12 subjects per condition. None of the subjects had participated in either Experiment 1 or 2.

Materials and procedure. The only changes made in Experiment 3 were the omission of the single-sentence context condition and the rearrangement of the booklets. Each booklet contained 114 pages, and each vocabulary word was presented three times. In the booklets used in the massed conditions, the word appeared three times in succession, whereas in the booklets used in the spaced conditions, there were 37 pages between each appearance of a word, just as in Experiment 2. In the massed conditions, therefore, the effective lag was 0, whereas in the spaced conditions, it was 37 (i.e., each appearance of a word was separated by 37 other words, or 4 min 19 s).

Results and Discussion

The scoring procedure used in Experiments 1 and 2 was used here also. The mean number of definitions correct was 13.42 for no-context control massed, 22.17 for no-context control spaced, 12.92 for three-sentence context massed, and 17.42 for three-sentence context spaced. A two-way ANOVA

applied to these data, with alpha set at the .05 level, revealed a reliable effect of lag (0 versus 37), $F(1, 44) = 6.57$, $p < .025$, and a nonsignificant effect of encoding condition (no context versus three-sentence context), $F(1, 44) = 1.03$. The Lag × Encoding Condition interaction failed to approach significance, $F(1, 44) < 1$, $MS_e = 80.15$.

As in Experiments 1 and 2, the number of intrusion errors (six) was too small to be of any significance. The number of misplaced definitions, however, did vary considerably across conditions—5 for no-context control massed, 7 for no-context control spaced, 23 for three-sentence context massed, and 14 for three-sentence context spaced. Thus, there were many more misplaced definitions in the three-sentence context conditions than in the no-context control conditions. But because the frequency of misplaced definitions did not covary systematically with the mean number of correct definitions, there does not appear to be any satisfactory explanation of these differences.

One potentially serious problem of interpretation arises from the fact that the retention interval for all but the final vocabulary word presented was confounded with presentation condition. Thus, it is possible that the observed spacing effect was due to the simple fact that subjects in the massed conditions had a longer period over which to forget the material than did subjects in the spaced conditions. Thus, it seemed prudent to conduct a fourth experiment, in which the retention interval for at least some of the vocabulary words was not confounded with spacing condition. Accordingly, I conducted a fourth experiment, using the conditions of Experiment 3 but having this objective in mind.

Experiment 4

Method

Subjects. Subjects were 48 students enrolled in undergraduate educational psychology courses at a state university and received extra course credit for their participation. They were randomly assigned to the four conditions, resulting in 12 subjects per condition. None of the subjects had participated in any of the previous experiments.

Materials and procedure. The only change made in Experiment 4 was in the arrangement of the booklets in the spaced conditions. The booklets were arranged so that the final repetition of 13 vocabulary words occupied the same pages in the two spacing conditions. Because of the constraints imposed by the manipulation of spacing condition, however, the retention interval for the final repetition of the remaining words continued to be confounded with spacing condition. The critical words in Experiment 4 were the final 13 words in Experiment 3.

Results and Discussion

The scoring procedure used in Experiments 1 through 3 was used. For the 13 critical vocabulary words, the mean number of definitions correct was 7.00 for no-context control massed, 8.58 for no-context control spaced, 4.58 for three-sentence context massed, and 9.25 for three-sentence context spaced. A two-way ANOVA applied to these data, with alpha set at the .05 level, revealed a significant effect of lag (0 versus 37), $F(1, 44) = 12.51$, $p < .01$, and a nonsignificant effect of encoding condition (no-context versus three-sentence context) $F(1, 44) < 1$. The Lag × Encoding condition interaction was not significant, $F(1, 44) = 3.05$, $MS_e = 9.37$.

For the 25 noncritical vocabulary words, the mean number of definitions correct was 12.25 for no-context control massed, 17.50 for no-context control spaced, 8.67 for three-sentence context massed, and 17.33 for three-sentence context spaced. A two-way ANOVA applied to these data, with alpha set at .05, revealed a reliable effect of lag, $F(1, 44) = 19.36$, $p < .001$, and a nonsignificant effect of encoding condition, $F(1,44) = 1.41$. The interaction also was not significant, $F(1, 44) = 1.67$, $MS_e = 30.01$.

As in the other experiments, the number of intrusion errors was too small to be of any significance, and the number of misplaced definitions, although larger, showed no systematic trend.

In conclusion, the results of the present experiment are generally consistent with those of Experiment 3 and support the hypothesis that spaced presentations lead to better vocabulary learning than do massed presentations. The only notable difference in the results of these two experiments is that in Experiment 3, the spacing effect was more apparent when words were accompanied only by their definitions, whereas in Experiment 4, the spacing effect was more apparent when words were also accompanied by sentences. In both experiments, however, the magnitude of the spacing effect was rather large. In Experiment 3, spaced presentations resulted in a 65% improvement when the words were accompanied only by their definitions and a 35% improvement when the words were also used in three different sentences. In Experiment 4, these improvements amounted to gains of 23% and 102%, respectively, for critical words and 43% and 100%, respectively, for noncritical words.

As for the issue of the confound in Experiment 3, the results for the critical words provide evidence that the spacing effect was not an artifact of differences in retention interval because they also yielded a spacing effect. Moreover, the recall probabilities associated with critical and noncritical words were nearly identical (57% and 56%, respectively), suggesting that differences in retention interval had a negligible effect on performance.

By contrast, the results of this experiment are consistent with the results of Experiments 1 through 3 in failing to support the hypothesis that multiple-sentence contexts lead to better vocabulary learning than does no context. As in the first three experiments, in fact, the trend was in the opposite direction.

Experiment 5

Even though McDaniel and Pressley (1984) found that the construction of adequate sentences including vocabulary items was determined largely by whether subjects knew the meanings of the items, a critic might charge that the present dependent variable is biased in favor of the definitions-only condition. Thus, it might be argued that had a context-related measure been selected, the results might have favored the

context conditions. Accordingly, I conducted a fifth experiment, in which I made some changes in the testing procedure. First, subjects were instructed that if they did not know the meaning of a word, they should attempt to write a sentence in which the word was used. Second, I constructed a totally new dependent measure, in which subjects were cued for the target words by sentence frames similar to those provided in the context conditions.

In addition, Experiment 5 included a completely new no-presentation baseline condition, in which subjects who were not exposed to the words and their definitions were simply administered the vocabulary tests. I included this condition to estimate the extent to which the present manipulations were effective in producing new learning. It is possible, of course, that the failure to obtain context effects in the present experiments was due to the way that encoding was manipulated, that is, a way that failed to produce much learning. Because little seemed to be gained by another comparison of massed versus spaced presentation, I eliminated the massed conditions, leaving only a no-context control spaced condition and a three-sentence context spaced condition.

Method

Subjects. Subjects were 36 students from the same classes that participated in Experiments 1 through 4 and received extra credit for their participation. They were randomly assigned to the three conditions, resulting in 12 subjects per condition. None of the subjects had participated in any of the previous experiments.

Materials and procedure. The only changes made in Experiment 5 relative to Experiment 4 were in the testing procedure and in the inclusion of the no-presentation baseline condition. The first test that subjects received was the sentence-cued recall test of 19 of the original 38 words. For this test, subjects were given a 2-page booklet with 19 randomly ordered sentences, each with a blank substituted for a particular target word. Subjects were instructed, "Attempt to fill in the blank in each of the sentences below with one of the words you studied." The sentences were similar in length and complexity to those used in the context conditions except that they did not contain any of the same key words. In addition, the sentences were written with the expectation that most subjects would know the meaning of the missing word. For example, "'Hitler's Diary,' after much publicity, was determined to be '———.'" The missing word was APOCRYPHAL. Following this test, subjects received test booklets that were the same as the ones used in the previous experiments except that subjects were instructed, "If you can't *define* the word, try to write a sentence using the word." The time limit for each test was 10 min.

Results and Discussion

For sentence-cued recall, a lenient spelling criterion was used. If the spelling of a correctly placed vocabulary word was at least a reasonable approximation of the correct spelling, credit was given. For the definitions-plus-sentences test, the scoring procedure used in Experiments 1 through 4 was used, but subjects were also given credit for any sentence indicating that they understood the meaning of the word in question. Not surprisingly, given McDaniel and Pressley's (1984) findings, subjects wrote very few sentences (15), and only 5 of those were acceptable. The sentence-cued recall and the definitions-plus-sentences recall means, respectively, were 0.25 and 2.67 for no-presentation baseline, 5.50 and 22.75 for no-context control, and 3.50 and 16.00 for three-sentence context.

The most striking feature of these results is that the no-presentation group performed so poorly. In fact, about half the subjects in this condition had a score of 0, and 1 subject accounted for 14 of the 32 definitions plus sentences correct. As might be expected then, I rejected the homogeneity of variance assumption for both measures, both Fs max(3, 11) > 5.00, $p < .05$, and did not conduct planned one-way ANOVAS with three levels. Instead, I subjected the scores of the no-context condition and the three-sentence context condition, which had comparable error variances, to t tests (two-tailed). With alpha set at .05, neither the differences in sentence-cued recall, $t(22) = 1.78$, nor the differences on the definitions-plus-sentences measure, $t(22) = 1.89$, were significant.

As has been the case with the definitions measure of learning, the vast majority of errors in sentence-cued recall and the definitions-plus-sentences measure were errors of omission. Furthermore, there were also few misplaced response errors. By contrast, subjects in the no-presentation condition often filled in blanks with words that though incorrect, indicated that they knew the meaning of the missing word. For example, 9 of the 12 subjects in this condition filled in the blank in the "Hitler's Diary" sentence with "phony," "falsified," "illagitamate [sic]," "untrue," "fake," "unauthentic," or "fraudulent." Thus, as planned, the meanings of the missing words tended to be fairly apparent.

Finally, there was a striking correspondence between the two measures in terms of the relative performance in the three conditions. The measures were highly redundant, as evidenced by the fact that 90% of the time, subjects who filled in a blank with the proper vocabulary word were also able to define it correctly. Apparently, the ability to use a word correctly in a ready-made context is highly dependent on the ability to define it.

In sum, the inclusion of a definitions-plus-sentences measure and a sentence-frame-cued recall measure in Experiment 5 did not reverse the results of the previous experiments with respect to the effectiveness of context. Instead, the no-context control group, even though not provided with any experience with vocabulary in context, performed better than the three-sentence context group on both measures. In addition, the difference in performance between the no-presentation condition and the encoding conditions indicates that the way in which encoding was manipulated in the present study was highly effective in producing vocabulary learning.

General Discussion

The results of the five experiments are summarized in Table 3. As can be seen, recall was not significantly affected by manipulations designed to affect the number of retrieval routes to the to-be-learned information. In particular, the three-sentence context condition failed to lead to better recall than did the single-sentence context condition, and both sentence conditions failed to lead to better recall than did the no-context control condition. Contrary to what would be expected from the encoding variability hypothesis, therefore,

Table 3
Recall Means and Standard Deviations for Experiments

Condition	Experiment 1	2	3	4	5 Definitions plus sentences	5 Sentence-cued recall[a]
Control						
Single presentation						
M	17.00					
SD	8.75					
Massed						
M			13.42	19.25		
SD			9.18	8.43		
Spaced						
M		19.17	22.17	26.08	22.75	5.50
SD		9.53	10.43	8.85	9.30	2.50
Definition + one sentence						
Single presentation						
M	15.33					
SD	10.36					
Massed						
M						
SD						
Spaced						
M		21.25				
SD		9.62				
Definition + three sentences						
Single presentation						
M	16.50					
SD	9.48					
Massed						
M			12.92	13.25		
SD			8.36	6.98		
Spaced						
M		18.83	17.42	26.58	16.00	3.50
SD		7.55	7.60	7.28	8.15	2.97
No-presentation baseline						
M					2.67	0.25
SD					4.01	0.45

[a] Maximum correct = 19.

the probability of recall did not vary directly with the number of potential retrieval routes to the target information. Although the power of each experiment was rather low, this conclusion is supported unequivocally by the results of all five experiments, and for the most part, the obtained differences were not even in the appropriate direction. By contrast, the results of Experiments 3 and 4 converge on the generalization that spaced presentations promote substantially higher levels of vocabulary learning than do massed presentations.

In view of several recent failures to support the encoding variability hypothesis (see Postman & Knecht, 1983), yet another failure may not be considered surprising to students of the learning and memory literature. The present failure is particularly damaging, however, inasmuch as the sentences provided a rich semantic context within which to associate the words and their meanings. Hence, the contextual material used in the present study should have permitted a particularly strong test of the encoding variability hypothesis.

On the other hand, the failure of sentence contexts to promote vocabulary learning relative to the no-context condition may well be unsettling to the many advocates of the learning-from-context approach to vocabulary learning. As indicated earlier, a critical assumption underlying this approach is that the appearance of a word in one or more sentences facilitates acquisition of its meaning. Indeed, as Sternberg et al. (1983) noted, it is from its emphasis on the interrelation of the word to the conceptual environment provided by sentences surrounding it that the learning-from-context method takes its name. This method, they suggested, is preferable to one in which the individual is exposed to a word and its definition.

In defense of the learning-from-context approach to vocabulary learning, several findings should be noted. First, a comparison between the performance of the no-presentation baseline condition and the sentence context conditions in the present study should make it clear that presenting words and their definitions in context is preferable to not presenting vocabulary words at all. Second, a recent set of studies in which students read sentences or passages containing unfamiliar words has shown that even in the absence of definitions, some learning from context does take place (Jenkins, Stein, & Wysocki, 1984; McKeown, 1985; Nagy, Herman, & Anderson, 1985). Third, studies by Christ and Petrone (1977), Gipe (1979),[2] and McKeown, Beck, Omanson, and Pople

[2] According to Pressley, Levin, Kuiper, Bryant, and Michener (1982), Gipe was subsequently unable to replicate the modest positive effects associated with her context method.

(1985) have demonstrated that some kinds of contextual enrichment can be more effective than simply presenting words and their definitions.

Thus, the results of the present study should not be interpreted to mean that context is of little or no value in the course of vocabulary learning. The present findings, together with those of McDaniel and Pressley (1984), do, however, indicate that the addition of context is not necessarily the most effective means of promoting vocabulary learning, even when there is a reasonable match between the contextual manipulation and the dependent variable, as there was in Experiment 5. Further, the present results call into question the assumption that multiple sentence contexts are better than one. The present results cannot simply be dismissed with Sternberg et al.'s (1983) warning that the addition of sentences may, under certain conditions, result in information overload that would interfere with the effective use of sentences. If that were the case, the single-sentence context groups should have clearly outperformed the three-sentence context groups, and obviously, I did not find this.

By contrast, the present results extend Dempster's (1986) and Reynolds and Glaser's (1964) in showing a robust effect of spacing in a task with clear classroom analogues. In addition, they suggest that previous failures to obtain significant spacing effects may have been due to the use of only relatively lengthy spacing intervals, 24 hr or more (Ausubel, 1966; Gay, 1973, Experiment 2; Peterson et al., 1935; Sones & Stroud, 1940). Strong spacing effects have been found only in studies using at least one condition in which the repetitions occurred in succession or no more than a few minutes apart. Thus, the principal practical implication of the spacing effect is that the teacher should avoid back-to-back or closely spaced repetitions.

The spacing effect is just one of a family of similar phenomena, all of which have implications for the distribution of time in the classroom. One such phenomenon is the "test-spacing" effect, which refers to the fact that spaced tests, particularly tests with intertest intervals of an expanding nature, result in greater retention than do massed testings (Landauer & Bjork, 1978; Rea & Modigliani, 1985). In one study, the retention of multiplication facts tested at expanding intervals was almost twice that found in a massed testing condition (Rea & Modigliani,1985).

Another related phenomenon has been observed when once-presented written exercises or a short course in statistics either is spread over the course of several sessions or is presented in a single session, a situation analogous to cramming for a test. In this case, students retain more when the material is distributed over several sessions (Bloom & Shuell, 1981; Smith & Rothkopf, 1984). Finally, Reder and Anderson (1982) found that with total study time equated, repeated and well-spaced presentations of a portion of text were more effective than was a single but longer presentation. In short, the spacing effect and allied phenomena appear to have reasonably clear implications for improving the management of time in the classroom.

Finally, the results of the present experiments bear directly on two theoretical issues. First, is recall more effective when multiple retrieval routes or a single retrieval route is established to the to-be-learned information? As Postman and Knecht (1983) noted, increasing the number of retrieval routes can, indeed, enhance the probability of recall. But, there are also good reasons to believe "that the efficacy of any given cue does increase with repeated processing of the cue–target relation" (p. 134). Thus, it is by no means obvious what predictions are to be made about the relative effectiveness of single and multiple retrieval routes. In the present case, the rather small and nonsignificant differences in recall found among the conditions of encoding suggest a trade-off between the strength of a retrieval route, which would be expected to increase with repeated processing, and the number of alternative retrieval routes. In short, the present results clearly demonstrate that the opportunity to establish multiple retrieval routes is not, in Postman and Knecht's (1983) words, "a sufficient condition of improved recall" (p. 133).

Second, the results of the present experiments drive yet another wedge between encoding variability theory and its credibility as an explanation of the spacing effect. In this case, though, the findings seem especially damaging for the simple reason that whereas spaced presentations resulted in significantly better recall than did massed presentations, there was no independent effect of the encoding manipulation. Accordingly, the observed effects of spacing must be attributable to some other mechanism. Although I did not design the present experiments to evaluate alternative models of the spacing effect, the total lack of support for the encoding variability account, ipso facto, heightens the credibility of other accounts, such as the inattention hypothesis, its principal competitor. In general, according to the inattention hypothesis, massed presentations receive less attention than do spaced presentations, as a result of either boredom or some sort of habituation process operating under massed conditions (Hintzman, 1974). Either or both of these processes could have been operating in the present study.

References

Ausubel, D. P. (1966). Early versus delayed review in meaningful learning. *Psychology in the Schools, 3*, 195–198.

Bjork, R. A. (1979). Information-processing analysis of college teaching. *Educational Psychologist, 14*, 15–23.

Bloom, K. F., & Shuell, T. J. (1981). Effects of massed and distributed practice on the learning and retention of second-language vocabulary. *Journal of Educational Research, 74*, 245–248.

Bower, G. H. (1972). Stimulus-sampling theory of encoding variability. In A. W. Melton & E. Martin (Eds.), *Coding processes in human memory* (pp. 85–124). New York: Winston & Sons.

Christ, R. L., & Petrone, J. M. (1977). Learning concepts from contexts and definitions. *Journal of Reading Behavior, 9*, 301–303.

Clifford, M. M. (1981). *Practicing educational psychology.* Boston: Houghton Mifflin.

Dempster, F. N. (1986). *Spacing effects in text recall: An extrapolation from the laboratory to the classroom.* Manuscript submitted for publication.

English, H. B., Wellborn, E. L., & Killian, C. D. (1934). Studies in substance memorization. *Journal of General Psychology, 11*, 233–260.

Gay, L. R. (1973). Temporal position of reviews and its effect on the retention of mathematical rules. *Journal of Educational Psychology, 64*, 171–182.

Gipe, J. (1979). Investigating techniques for teaching word meanings. *Reading Research Quarterly, 14*, 624–644.

Glenberg, A. M. (1977). Influences of retrieval processes on the spacing effect in free recall. *Journal of Experimental Psychology: Human Learning and Memory, 3*, 282–294.

Hintzman, D. L. (1974). Theoretical implications of the spacing effect. In R. L. Solso (Ed.), *Theories in cognitive psychology: The Loyola Symposium* (pp. 77–99). Potomac, MD: Erlbaum.

Jenkins, J. R., Stein, M. L., & Wysocki, K. (1984). Learning vocabulary through reading. *American Educational Research Journal, 21*, 767–787.

Johnson, D. D., & Pearson, P. D. (1984). *Teaching reading vocabulary* (2nd ed.). New York: Holt, Rinehart & Winston.

Landauer, T. K., & Bjork, R. A. (1978). Optimum rehearsal patterns and name learning. In M. M. Gruneberg, P. E. Morris, & R. N. Sykes (Eds.), *Practical aspects of memory* (pp. 625–632). New York: Academic Press.

Martin, E. (1972). Stimulus encoding in learning and transfer. In A. W. Melton & E. Martin (Eds.), *Coding processes in human memory* (pp. 59–84). New York: Winston & Sons.

McDaniel, M. A., & Pressley, M. (1984). Putting the keyword method in context. *Journal of Educational Psychology, 76*, 598–609.

McKeown, M. G. (1985). The acquisition of word meaning from context by children of high and low ability. *Reading Research Quarterly, 20*, 482–496.

McKeown, M. G., Beck, I. L., Omanson, R. C., & Pople, M. T. (1985). Some effects of the nature and frequency of vocabulary instruction on the knowledge and use of words. *Reading Research Quarterly, 20*, 522–535.

Melton, A. W. (1970). The situation with respect to the spacing of repetitions and memory. *Journal of Verbal Learning and Verbal Behavior, 9*, 596–606.

Nagy, W. E., Herman, P. A., & Anderson, R. C. (1985). Learning words from context. *Reading Research Quarterly, 20*, 233–253.

Peterson, H. A., Ellis, M., Toohill, N., & Kloess, P. (1935). Some measurements of the effects of reviews. *Journal of Educational Psychology, 26*, 65–72.

Postman, L., & Knecht, K. (1983). Encoding variability in retention. *Journal of Verbal Learning and Verbal Behavior, 22*, 133–152.

Pressley, M., Levin, J. R., Kuiper, N. A., Bryant, S. L., & Michener, S. (1982). Mnemonic versus nonmnemonic vocabulary-learning strategies: Additional comparisons. *Journal of Educational Psychology, 74*, 693–707.

Rea, C. P., & Modigliani, V. (1985). The effect of expanded versus massed practice on the retention of multiplication facts and spelling lists. *Human Learning, 4*, 11–18.

Reder, L. M., & Anderson, J. R. (1982). Effects of spacing and embellishment on memory for the main points of a text. *Memory & Cognition, 10*, 97–102.

Reynolds, J. H., & Glaser, R. (1964). Effects of repetition and spaced review upon retention of a complex learning task. *Journal of Educational Psychology, 55*, 297–308.

Smith, S. M., & Rothkopf, E. Z. (1984). Contextual enrichment and distribution of practice in the classroom. *Cognition and Instruction, 1*, 341–358.

Sones, A. M., & Stroud, J. B. (1940). Review with special reference to temporal position. *Journal of Educational Psychology, 31*, 665–676.

Sternberg, R. J., Powell, J. S., & Kaye, D. B. (1983). Teaching vocabulary-building skills: A contextual approach. In A. C. Wilkinson (Ed.), *Classroom computers and cognitive science* (pp. 122–143). New York: Academic Press.

Toppino, T. C., & Gracen, T. F. (1985). The lag effect and differential organization theory. *Journal of Experimental Psychology: Learning, Memory and Cognition, 11*, 185–191.

Received December 16, 1985
Revision received October 1, 1986
Accepted November 5, 1986 ■