

# Word recall as a function of sentence generation and sentence context

DAN GOLLUB and ALICE F. HEALY  
University of Colorado, Boulder, Colorado

Subjects were asked to generate long and short sentences using target words, to estimate the effort involved in producing each sentence, and then to recall the target words. Yoked subjects were asked to evaluate target word usage in the sentences generated by the first group and then were assigned the same recall task. Target word recall was significantly greater for subjects generating than for subjects evaluating sentences and for long as opposed to short sentences. Also target word recall was significantly correlated with evaluation ratings but not with effort ratings. The implications of these findings are discussed with respect to theoretical accounts of the generation effect and to practical concerns of educators in teaching new vocabulary items.

Numerous investigators have found evidence that people who generate verbal material tend to remember it better than people who only read that material (e.g., Slamecka & Graf, 1978). This phenomenon has been called the *generation effect*. One possible explanation for it is that the cognitive effort involved in generating the material assists in its subsequent recall (e.g., McFarland, Frey, & Rhodes, 1980; Tyler, Hertel, McCallum, & Ellis, 1979). The difficulties involved in quantifying the amount of cognitive effort expended, however, has proven a barrier to delineating precise relationships between effort and recall.

Another factor affecting recall for verbal material is the context in which it appears. For example, Craik and Tulving (1975) found that subjects tended to remember words that appeared in complex sentences better than words that appeared in simple sentences.

The present experiment was designed to supplement both of these lines of investigation. Subjects were given a list of 30 target words and were asked to generate either long (complex) or short (simple) sentences from each word. In order to provide an assessment of cognitive effort, the subjects also were asked to estimate (on a 7-point scale) how difficult each sentence was to generate. Following completion of these tasks, they were given the unexpected task to recall as many of the target words as they could. A yoked group of subjects was asked to evaluate the sentences that had been produced, and then were given the same unexpected recall task. It was predicted that the subjects who generated the sentences would remember

more target words than the subjects who evaluated those sentences, that subjects who generated the sentences would recall more words from sentences rated as difficult than from those rated as easy, and that all subjects would remember more words from long than from short sentences.

## METHOD

### Subjects

The subjects were 32 male and female undergraduate psychology students who volunteered for the study to earn course credit. There were two conditions (generate and evaluate) with 16 subjects in each condition. Subjects in the generate condition were divided into two subgroups ( $n=8$  in each subgroup). Subjects were assigned to conditions and subgroups solely on the basis of order of arrival for testing, with those in the evaluate condition tested after those in the generate condition.

### Materials

Subjects were provided with a list of the 30 target words. Those words, all of which were nouns, were taken from a set of norms produced by Paivio, Yuille, and Madigan (1968) and were chosen to include a variety of levels of word frequency, abstractness, and concreteness. The order of the words was alphabetical.

### Procedure

Subjects were tested in groups of 1 to 3 individuals. Those in one generate subgroup were asked to generate long sentences (eight words or more) from the odd-numbered target words and short sentences (six words or fewer) from the even-numbered target words; subjects in the other generate subgroup were asked to generate short sentences from the odd-numbered target words and long sentences from the even-numbered target words. (The experimenter monitored the subjects' responses to ensure that the sentences were of the desired length.) All subjects in the generate condition were asked to estimate on a 7-point scale, immediately after producing each sentence, the difficulty of generating that sentence (1 = *easy*, 7 = *most difficult*).

Each subject in the evaluate condition was provided the 30 target words along with the corresponding sentences produced by a yoked subject from the generate condition, and was asked to evaluate on a 7-point scale how well each sentence used the target word (1 = *poor*, 7 = *excellent usage*).

Following completion of the preliminary tasks, all subjects were given a surprise recall task: They were asked to write down as many of the 30 target words as they could remember. The subjects were informed that they would have 5 min for this task.

---

This research was supported in part by Contract MDA903-86-K-0155 from the United States Army Research Institute to the Institute of Cognitive Science at the University of Colorado, Boulder. We are indebted to Lyle Bourne, Edward Crothers, and Robert Crutcher for helpful discussions of this research and comments on an earlier draft of this article.

Requests for reprints should be addressed to Alice F. Healy, Department of Psychology, University of Colorado, Campus Box 345, Boulder, CO 80309.

**Table 1**  
**Mean Number of Correct Word Recalls (out of 15) as a Function of Condition and Sentence Length**

Condition	Length		Combined
	Short	Long	
Generate	5.75	7.06	6.41
Evaluate	3.37	4.81	4.09
Combined	4.56	5.94	

## RESULTS

Table 1 summarizes the results in terms of the mean number of correct recalls as a function of condition and sentence length (with 15 a perfect score for each condition). Generation did yield better recall performance than did evaluation, as expected. Word recall means were 6.41 for generate subjects versus 4.09 for evaluate subjects [ $F(1,15) = 13.2$ ,  $MSe = 6.46$ ,  $p < .01$ ]. In addition, as anticipated, the long sentences yielded better recall performance than did the shorter sentences. Word recall means were 5.94 for target words in long sentences versus 4.56 for target words in short sentences [ $F(1,15) = 7.2$ ,  $MSe = 4.22$ ,  $p < .05$ ]. The interaction of condition and sentence length was not significant [ $F(1,15) < 1$ ].

Effort ratings by the generate subjects were not influenced by sentence length. The average effort rating for long sentences was 2.85, and the average effort rating for short sentences was 2.68 [ $F(1,15) < 1$ ]. In contrast, sentence evaluation ratings by the evaluate subjects did vary as a function of sentence length. Average evaluation of long sentences was 4.56 and that of short sentences was 4.13 [ $F(1,15) = 4.48$ ,  $MSe = 78.39$ ,  $p < .05$ ].

The Pearson product-moment correlation coefficient of average effort rating for the sentences in which a word appeared with the number of times that word was recalled by subjects in the generate condition was only .24 ( $p > .05$ ). The Pearson product-moment correlation coefficient of mean evaluative ratings (across subjects) for the sentences in which a target word appeared with the number of times that word was recalled by subjects in the evaluate condition was .36 ( $p = .05$ ). Surprisingly, the effort ratings associated with a given target word were negatively correlated with the evaluations of the sentences in which those target words appeared ( $r = -.64$ ,  $p < .01$ ).

## DISCUSSION

The significantly better performance by generate subjects than by evaluate subjects lends support to previous findings concerning the generation effect. Thus, the statement by Slamecka and Graf (1978) that the generation effect manifests itself across a variety of testing procedures, encoding rules, and other situational changes was confirmed. Similarly, the findings by Craik and Tulving (1975) regarding the effects of sentence context complexity on recall were also supported. The findings in this experiment that effort ratings for a sentence were not a function of sentence length and did not significantly correlate with recall are not seen as inconsistent with past investigations of the generation effect. For example, Zacks, Hasher, Sanft, and Rose (1983) queried whether amount of effort expended was a principal predictor of recall, and McElroy and Slamecka (1982) argued against an effort explanation of

the generation effect. Furthermore, McFarland et al. (1980) acknowledged that the generation effect occurs even for low-level cognitive tasks, which might not require much effort.

The surprising negative correlation between evaluative ratings and effort ratings similarly indicates that effort expended is not necessarily a significant predictor of performance. A possible explanation for that negative correlation is that difficult sentences required more effort, but that effort did not entirely compensate for the difficulty involved. Thus poorer sentences resulted.

Recalling a set of items has frequently been described as involving an associative search (e.g., Anderson, 1972). Assuming this to be so, the findings that subjects evaluated target word usage as being better in long than in short sentences and that the long sentences yielded better recall performance than did the short sentences suggest that more associations were created in the long sentences. Thus, the better recall is attributed to the larger number of associations created within the long sentences. However, it is hypothesized that it is not the mere quantity of words that produces associations and thereby enhances recall. The significant positive correlation found in this experiment between sentence evaluation and recall of target words suggests that short sentences, if well written, can also yield associations helpful in recall. A similar conclusion was arrived at by Stein, Morris, and Bransford (1978), who found that effective semantic elaboration cannot be equated with the mere quantity of semantically congruous information.

Sentence generation is one of the tools that educators use to facilitate vocabulary development (Beck & McKeown, 1983). The finding in this experiment that long sentences were not significantly more difficult to generate than short sentences, in conjunction with the finding that subjects remembered target words from long sentences significantly better than target words from short sentences, leads to the conclusion that when sentence generation is employed to assist in the teaching of new words, long rather than short sentences should be used. The associations produced in the long sentences should aid in the subsequent retrieval of the new words.

## REFERENCES

- ANDERSON, J. R. (1972). FRAN: A simulation model of free recall. In G. H. Bower (Ed.), *The psychology of learning and motivation* (Vol. 5, pp. 315-378). New York: Academic Press.
- BECK, I. L., & MCKEOWN, M. G. (1983). Learning words well: A program to enhance vocabulary and comprehension. *Reading Teacher*, 36, 622-625.
- CRAIK, F. I. M., & TULVING, E. (1975). Depth of processing and the retention of words in episodic memory. *Journal of Experimental Psychology: General*, 104, 268-294.
- MCELROY, L. A., & SLAMECKA, N. J. (1982). Memorial consequences of generating nonwords: Implications for semantic-memory interpretations of the generation effect. *Journal of Verbal Learning & Verbal Behavior*, 21, 249-259.
- MCFARLAND, C. E., JR., FREY, T. J., & RHODES, D. D. (1980). Retrieval of internally versus externally generated words in episodic memory. *Journal of Verbal Learning & Verbal Behavior*, 19, 210-225.
- PAIVIO, A., YUILLE, J. C., & MADIGAN, S. A. (1968). Concreteness, imagery, and meaningfulness values for 925 nouns. *Journal of Experimental Psychology, Monograph Supplement*, 76(1, Pt. 2).
- SLAMECKA, N. J., & GRAF, P. (1978). The generation effect: Delineation of a phenomenon. *Journal of Experimental Psychology: Human Learning & Memory*, 4, 592-604.
- STEIN, B. S., MORRIS, C. D., & BRANSFORD, J. D. (1978). Constraints on effective elaboration. *Journal of Verbal Learning & Verbal Behavior*, 17, 707-714.
- TYLER, S. W., HERTEL, P. T., MCCALLUM, M. C., & ELLIS, H. C. (1979). Cognitive effort and memory. *Journal of Experimental Psychology: Human Learning & Memory*, 5, 607-617.
- ZACKS, R. T., HASHER, L., SANFT, H., & ROSE, K. C. (1983). Encoding effort and recall: A cautionary note. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, 9, 747-756.