

Delayed Retrieval Following Text Synthesis with Varied Feedback

Verne Keenan

Philip Langer

Christina M. Medosch-Schonbeck

University of Colorado

Technical Report #136

Institute of Cognitive Science

University of Colorado

Boulder, Colorado 80309

February, 1985

Delayed Retrieval Following Text Synthesis with Varied Feedback

Verne Keenan, Philip Langer and Christina M. Medosch-Schonbeck

That information feedback facilitates performance and learning is generally accepted by educators and psychologists, but the mechanisms of that facilitation for particular tasks are seldom elaborated and are insufficiently understood. The improvement of instruction through improved use of feedback requires detailed knowledge of specific applications in particular contexts. Toward this goal of finer specification, Langer and Keenan (1984) have begun the exploration of feedback during the synthesis of scrambled text, hoping to reveal components of the ongoing construction of meaning from written discourse. Subjects were asked to construct sensible passages from randomly ordered sentences. The present investigation extends the previous study to include longer passages and to probe the influence of text synthesis on delayed retrieval.

Concern for the appropriate accommodation of feedback procedures to target tasks grows out of the too common overgeneralization of the meanings of the terms feedback and learning. Perhaps the most common of these generalizations made by several instructional psychologists, is the failure to differentiate feedback from reinforcement. Skinner, for example, in three decades, seems not to have modified his ideas about teaching and school learning (see Skinner, 1954, 1984). From this operant perspective, the learner makes an overt response with a high likelihood of success and is reinforced by information feedback which is positive and immediate. These three behavior-shaping principles may serve well when applied to routine classroom behaviors such as putting the lid back on the glue, reinforced by praise or a hug from the kindergarten teacher. Most educational psychologists had learned by the end of the 1960's, however, that behavior-

shaping principles in programmed instruction materials are not generally suited to all curricular tasks, and counterexamples are easy to find (see, e.g., Lumsdaine, 1963; Zoll, 1969). Immediacy of reinforcement is not required or even optimal for many school-like tasks (Surber & Anderson, 1975; Sturges, 1978). Further, it is not necessary that information feedback be positive and reinforcing to be useful in learning. The confirm-disconfirm dimension of instruction, and the information-incentive dimension, are clarified by Getsie, Langer, and Glass (1985) who have given a robust review of feedback experiments which indicates many of the important contextual parameters that should modulate our instructional applications. For discrimination problem-solving by children, they find confirmation less helpful than disconfirmation.

In light of this "disconfirming" empirical evidence, it may seem a bit surprising that feedback has not been completely abandoned in the design of instructional systems, but current examples of its use are easy to find. In a recent discussion of the behavioral approach to teaching and learning we find "...we have assumed that knowledge of results is sufficiently reinforcing... The student is provided with feedback as to the accuracy of his or her response," (Mason, Cegelka, Lewis, Henry, Larkin & Danner, 1983). And further, a major theory of instructional design includes the rule, "Instance-practice items or generality-practice items should be followed by immediate feedback as to whether....," (Merrill, Kowallis & Wilson, 1981). This persistence of feedback recommendations is apparently due to the fact that feedback during performance is often helpful, and sometimes critically necessary in particular cases of learning and instruction.

It is common sense, of course, that driving an automobile with eyes closed leads to disaster, and that composing a computer program is greatly facilitated by the ongoing feedback of a monitor. What needs to be done is

to tailor learning principles to instructional tasks. We must learn to partition instructional episodes into events which are sufficiently specified to permit accurate prediction of whether, and what kind of, informational guidance will enhance learning. In this way, a gradual assembly of such events with explicated varieties of information feedback, might provide custom-tuned instructional interventions. The general task under investigation here is the construction of meaning from expository written discourse. The goal is to find feedback-sensitive subtasks.

We have assumed that reading for information is a problem-solving task in which the reader attempts to establish correspondence between personal substantive and strategic knowledge and some message implicit in the text being read. It is further assumed that tentative interpretations are tested for coherence and correspondence during reading, and that such testing can be facilitated by information concerning the ongoing interpretive processing. The passages were chosen to contain familiar vocabulary in sentences which follow conventional syntax. The sentences were presented in random order to reveal, in part, the processes used by readers in making sense from the materials. The specific feedback strategy chosen was a provision of right-wrong information following certain sentence-order decisions.

In our prior investigation of feedback during text synthesis, the subjects were given the individual sentences of short texts typed on cards and randomly ordered. As they rearranged the cards to reconstruct the original passage, one group had access to confirm-disconfirm information and the other group did not. It was assumed that feedback would enhance the reconstruction of the original sequence and that subsequent retrieval of the text would be facilitated to the extent of the success of reconstruction. This approximation to the original was indexed by Kendall's tau coefficients

between each subject's final construction and the original. Although tau values were correlated moderately with recall scores, neither recall nor tau was reliably different across feedback and no-feedback groups. The ability these subjects showed, to recall nearly half (44%) of the idea units in written recall seemed not to depend on their understanding of the construction and meaning of the whole text. Some subjects mentioned West Germany or England rather than Germany as our competitor in the World War II weapons race, evidence of poor semantic integration. These subjects showed good recall, nonetheless, of related materials.

Other factors were passage length and idea density. Percent recall and percent recognition both favored shorter versions of more dense and complex passages. These anomalous results suggest that memory for text depends on surface features, situational factors or local clusters of ideas more than on the gist of the total text (see van Dijk & Kintsch, 1983; Young & Bourne, 1982). Feedback of information about sequence was intended to support overall understanding of the passages rather than superficial rote encoding. In order to make comprehension of the integral whole more salient, the present experiment uses a longer and denser (more ideas) version than in the prior experiment, and includes a delayed recall test to see if understanding gist enhances retrieval after a longer interval.

METHOD

Subjects and Design

Fifty-four volunteers from the Introductory Psychology subject pool served in the nine conditions of the experiment (36 from the Spring 1984 term and 18 from the Summer 1984 term). Three factors were crossed to make eight conditions: number of sentences (13 or 26), feedback or no-feedback, and immediate or one-week delayed retrieval. The ninth condition was a control which involved reading only rather than reconstruction processing.

Control subjects read the shorter passage twice, without feedback, and gave immediate retrieval.

Materials

As in our prior investigation, passages were adapted from the biography of Enrico Fermi written by his wife, Laura. The short selection was included in the longer selection describing a part of the work on the first man made atomic chain reaction which occurred in 1942 at the University of Chicago. Inspection of these texts (see Appendix) shows that they are in familiar, non-technical, language with exceptions such as (atomic) pile. Most readers would be able to read and draw tentative partial meanings from the individual sentences without knowing that the topic was The Manhattan Project of World War II. There were 56 idea units in the short version (4.3 per sentence) and 107 idea units in the longer version (4.1 per sentence). Sentences were typed on index cards, one per card, and arranged in an order generated from a random number table before being given to the subjects. A board with 35 slots was provided to display the cards. Subjects, individually, were required to read the cards, one at a time, and place each card in a slot in the board as soon as it was read; and then to rearrange the cards by moving one at a time until they were in the sequence that had the best order or made the best sense. Feedback subjects were given five tokens which could be offered in exchange for confirmation or disconfirmation at any move. A paraphrase was written for each sentence and paired, randomly ordered, with its original for a recognition test. A short practice task, the two passages, and the recognition tests appear in the Appendix.

Procedures

Subjects were randomly assigned to passage length and feedback treatment combinations as they appeared for the experiment. Following

treatment-specific instructions the practice task was explained and performed with the slot-board and a short deck about The Goose that Laid the Golden Egg. After any questions about procedures were resolved, the subject was given the experimental deck appropriate to the assigned treatment, i.e., a randomized 13- or 26-sentence deck for experimental subjects and the 13 sentences in original order for control subjects. Each subject in the feedback condition received five tokens which could be exchanged for feedback following a move, at the discretion of the subject. A move, for any subject, was counted when any card was taken from one slot and put in another. If a sentence was moved to the position immediately following that of the sentence which it followed in the original text it was called right. Any other move was wrong. The experimenter counted moves and processing time until the subject was satisfied that the most sensible sequence was attained. Half the subjects stayed for a written recall test and the old-new recognition test. The other half returned after a week to take the retrieval tests.

RESULTS AND DISCUSSION

Independent variables were number of sentences, presence or absence of feedback during text synthesis, and immediate or delayed retrieval. Dependent measures were recall, recognition and tau. Ancillary measures were text processing time, moves during processing, and number of tokens used (in feedback conditions).

Recall

Written recall protocols were scored for idea units following Kieras and Bovair (1981), and raw scores were transformed to percentages of the 56 ideas in the short passage or the 107 ideas in the long passage. Mean

recall percentages are shown (with standard deviations) in Table 1. As in the previous experiment, the shorter passage was recalled better than the longer passage, $F(1,40)=5.75$, $p<.05$.

 Insert Table 1 about here

Feedback had no effect, $F>1$. With the longer passage increased in length and density, relative to the first experiment, we hoped to see feedback effects with the more difficult passage, i.e., information should be more helpful when the task exceeds the capacity of a surface encoding and requires the cuing provided by assimilation to existing schemata. The immediate recall scores were higher than one-week delayed recall scores, $F(1,40)=9.79$, $p<.01$. Again, there was the expectation that feedback would be useful in establishing more durable associative structures, but the interaction (albeit in the expected pattern) was not statistically significant, $F(1,40)=2.87$, $p>.05$.

Recognition

The recognition test was scored for the number of correct "old" sentences identified, and scores were transformed to percentages of the 13 or 26 sentences processed. Mean recognition percentages (and standard deviations) appear in Table 2. Since the paraphrasing of sentences for the test did not change meaning, it might be expected that the organization provided by sequence information feedback would not affect recognition. It

 Insert Table 2 about here

is usually assumed that recognition is a relatively direct process in which an item is called "old" if it is familiar beyond some criterion (Kintsch,

1970), though organizational effects are reliably found with certain paradigms (e.g., Mandler, Pearlstone & Koopmans, 1969). Item discriminability also influences recognition (Juola, Fischler, Wood & Atkinson, 1971), and intersentence comparison would enhance discriminability. All factors showed differences. Recognition scores were higher for the shorter text, $F(1,40)=16.7$, $p<.01$; higher for feedback than no-feedback condition, $F(1,40)=17.0$, $p<.01$; and higher for immediate than delayed testing, $F(1,40)=13.2$, $p<.01$. Although some part of these effects may result from intrinsically high reliabilities of recognition scores, there is also the likelihood that subjects are encoding into rote episodic memory, the detailed surface features of sentences used in these selections; and little assimilation to meaningful schemata is involved. There is further evidence that the integrated meaning of the entire text is not the effective schema maintained in memory. The delay effect is considerably stronger for the 26-sentence than for the 13-sentence version, $F(1,40)=4.33$, $p<.05$. This is contrary to the notion that meaning should be richer in the longer version and minimize loss at delayed retrieval, but compatible with the idea of intersentence comparison.

Tau

Kendall's tau coefficients were calculated between the sentence order each subject produced and the order in the original passage. Mean coefficients are given in Table 3. Agreement with the original was higher

Insert Table 3 about here

for the feedback than for the no-feedback condition, $F(1,40)=5.04$, $p<.05$.

Also in Table 3 are the concordances (Kendall's w) for the eight cells.

These coefficients index the agreement of subjects with each other, rather

than agreement with an original order which they were asked to duplicate. The two sets of coefficients are in complete rank-order agreement for the eight cells. In addition, for each coefficient, the four feedback-no-feedback comparisons all favor the feedback condition. Like the lady tasting tea (Fisher, 1956), choosing correctly which four of the eight had feedback (or milk) would occur by chance only once in seventy tries.

What conclusions can be drawn from these results? Subjects made closer approximations to the original sequential structure of a passage when given right-wrong information during text synthesis. How general this outcome might be is quite unclear. In the prior, similar experiment, tau was only 20 percent greater for feedback subjects, a chance-level difference. Even if the difference is general across subjects and passages like these, it may not apply unless feedback is given only on request by the subject. If it is a general, valid effect, why does the concordance not lead to better recall? In the prior experiment, the Pearson correlation of tau with recall was .25, significant at the .05 level, but in the present data the coefficient is only .09, nonsignificant.

Certain other results suggest tentative inferences which may be drawn from these data. Tau is positively and reliably related to processing time and, for feedback subjects, to the number of tokens used. The evidence converges on the notion that feedback on demand influences some subjects to process text more thoroughly and to make a superior synthesis. Retrieval however appears not to be influenced by this kind of processing.

A final bit of evidence comes from the retrieval performances of the control subjects. There were four of these who read the 13-sentence version two times in the original sequence, then completed the same recall and recognition tests as the others. Their scores should be compared with others who processed the short version, had no feedback, and had no delay.

For this group and the controls, processing times were 7.85 min and 7.90 sec. Recall percentages were 35.17 and 34.25, and recognition percentages were 75.7 and 86.6. Apparently the text synthesis process is not more helpful, on balance, than simply reading the sentences as an encoding activity.

We know that artificial texts can be written so that the reader organizes an internal representation which depends on cues external to the presented text (e.g., Anderson & Pichert, 1978; Bransford & Johnson, 1973). Our attempt to avoid artificial texts has evidently introduced variables which conceal some of the processes we have assumed, and also some of our assumptions may be quite invalid. Choice of texts for these experiments was guided by a desire for a natural expository or narrative discourse with familiar language and conventional syntax but relatively unfamiliar content. Otherwise the choice was quite arbitrary, and the sequential dependence of the propositional content included in the Fermi texts was surprisingly weak. Discourse with tighter sequential logic should be found or written to probe parameters of the effects of external guidance.

As a final note on discourse processing, there is a need to consider what kinds of representations permit the recall of 30 or more ideas from scrambled text (seven subjects performed this well). Nonsense phrases would not support such success. There is a "local coherence" construct discussed by van Dijk and Kintsch (1983) which depends on argument overlap between and among propositions. In text synthesis, this argument overlap may be revealed in subjects' final sequence of sentences and again in recall protocols if they are taken with appropriate constraints. The amounts of artificiality of texts, of intrinsic argument overlap in texts, and of sequential clustering in recall need to be investigated in the context of the text synthesis and feedback procedures.

References

- Anderson, R.C., & Pichert, J.W. (1978). Recall of previously unrecalable information following a shift in perspective. Journal of Verbal Learning and Verbal Behavior, 17, 1-12.
- Bransford, J.D., & Johnson, M.K. (1973). Considerations of some problems of comprehension. In W.G. Chase (Ed.), Visual information processing. New York: Academic Press.
- Fisher, R.A. (1956). An example: Mathematics of a lady tasting tea. Reprinted in J.R. Newman, The world of mathematics, vol. III. New York: Simon and Shuster.
- Getsie, R., Langer, P., & Glass, G.V. (in press). Meta-analysis of the effects of type and combinations of feedback on children's discrimination learning. Review of Educational Research.
- Juola, J.F., Fischler, I., Wood, C.J., & Atkinson, R.C. (1971). Recognition time for information stored in long-term memory. Perception and Psychophysics, 10, 8-14.
- Kintsch, W. (1970). Models for free recall and recognition. In D. Norman (Ed.), Models of human memory. New York: Academic Press.
- Langer, P., & Keenan, V. (1984). Modification of current feedback strategies: A text synthesis approach. Technical Report No. 127, Institute of Cognitive Science, Boulder, Colorado 80309.
- Lumsdaine, A.A. (1963). Instruments and media of instruction. In N.L. Gage (Ed.), Handbook of research on teaching. Chicago: Rand McNally.
- Mandler, G., Pearlstone, Z., & Koopmans, H.J. (1969). Effects of organization and semantic similarity on recall and recognition. Journal of Verbal Learning and Verbal Behavior, 8, 410-423.

- Mason, E.J., Cegelka, P., Lewis, R., Henry, S., Larkin, J., & Danner, F. (1983). Three approaches to teaching and learning in education: Behavioral, Piagetian, and information-processing. Instructional Science, 12, 219-241.
- Skinner, B.F. (1954). The science of learning and the art of teaching. Harvard Educational Review, 24, 86-97.
- Skinner, B.F. (1984). The shame of American education. American Psychologist, 39, 947-954.
- Sturges, P.T. (1978). Delay of informative feedback in computer-assisted testing. Journal of Educational Psychology, 70, 280-288.
- Surber, J.R., & Anderson, R.C. (1975). Delay-retention effects in natural classroom settings. Journal of Educational Psychology, 67, 170-173.
- van Dijk, T.A. & Kintsch, W. (1983). Strategies of discourse processing. New York: Academic Press.
- Young, S.R. & Bourne, L.E., Jr. (1982). Idea organization and idea recall. Technical Report No. 119, Institute of Cognitive Science, University of Colorado, Boulder.
- Zoll, E.J. (1969). Research in programmed instruction in mathematics. Mathematics Teacher, 62, 103-110.

Appendix

This is the recognition test for the 26-sentence condition. Items of the 13-sentence version are numbered at the left. The correct original sentences have been marked with an asterisk.

- 1) 1. At the Met Lab, Herbert Anderson had been building small piles with his group and gathering information.

*Herbert Anderson and his group at the Met Lab had been building small piles and gathering information for a larger pile.

- 2) 2. The squash court under the West stands of the University stadium, Stagg Field, was the best place they had been able to find for work on the pile.

*The best place they had been able to find for work on the pile was a squash court under the West Stands of Stagg Field, the University stadium.

3. Stagg Field was used for odd purposes since President Hutchins had banned football from the Chicago campus.

*President Hutchins had banned football from the Chicago campus and Stagg Field was used for odd purposes.

4. The entrance to the space beneath the West stands is through a heavy portal on Ellis Avenue.

*On Ellis Avenue, through a heavy portal is the entrance to the space beneath the West Stands.

5. Part of this space, the Squash Court, was 30 feet by 60 feet and more than 26 feet in height.

*The Squash Court, which was part of this space, was 30 feet wide, twice as long, and over 26 feet high.

- 3) 6. Although the physicists wanted more space for the pile, the better places had been requisitioned by the expanding armed forces in Chicago.

*The physicists would have liked more space, but places better suited for the pile had been requisitioned by the expanding armed forces stationed in Chicago.

7. The physicists would have to be content with the Squash Court, where Herbert Anderson had started assembling piles.

*The physicists were to be contented with the Squash Court, and there Herbert Anderson had started assembling piles.

- 4) 8. The piles were still "small" because materials arrived at the West Stands at a very slow, if steady, pace.

*They were still "small piles," because material flowed to the West Stands at a very slow, if steady, pace.

- 5) 9. Herbert's spirits rose with the arrival of each new shipment of crates.

*As each new shipment of crates arrived, Herbert's spirits rose.

10. He loved work for he was of impatient temperament.

*He loved working and was of impatient temperament.

11. His body, though slender and almost delicate, had unsuspected endurance and resilience.

*His slender, almost delicate, body had unsuspected resilience and endurance.

- 6) 12. He worked at all hours and his same intensity and enthusiasm drove his associates to work along with him.

*He could work at all hours and drive his associates to work along with his same intensity and enthusiasm.

- 7) 13. The hired men who normally unpacked the crates were not working on the Saturday afternoon when the shipment arrived at the West Stands.

*A shipment of crates arrived at the West Stands on a Saturday afternoon when the hired men who normally unpacked them were not working.

- 8) 14. A university professor who was several years older than Herbert looked at the crates and said lightly, "Those fellows will unpack them Monday morning."

*A university professor, older by several years than Herbert, gave a look at the crates and said lightly, "Those fellows will unpack them Monday morning."

- 9) 15. "Those fellows, Hell!" flared Herbert, who never felt inhibited by the presence of men older and higher in the academic hierarchy, "We'll do them now."

*"Those fellows, Hell! We'll do them now," flared Herbert, who had never felt inhibited in the presence of older men, higher up in the academic hierarchy.

- 10) 16. As the professor took off his coat, the two men started wrenching at the crates.

*The professor took off his coat, and the two men started wrenching at the crates.

17. The profanity, which was used freely at the Met Lab, relieved the tension built up by working against time.

*Profanity was freely used at the Met Lab -- it relieved the tension built up by having to work against time.

11) 18. Would the Germans get atomic weapons before they were developed in the United States?

*Would Germany get atomic weapons before the United States developed them?

12) 19. Would these weapons be ready in time to help win the war?

*Would these weapons come in time to help win the war?

13) 20. Constantly present in the minds of the leaders of the project, these unanswered questions pressed them to work faster and faster and to be tense and to swear.

*These unanswered questions constantly present in the minds of the leaders of the project pressed them to work faster and faster, to be tense, and to swear.

21. By spring a small pile assembled in the Squash Court showed that all conditions were such that a pile of critical size would chain-react, and success was assured.

*Success was assured by the spring when a small pile assembled in the Squash Court showed that all conditions were such that a pile of critical size would chain-react.

22. "It could have been May or early June," said Enrico, "I remember talking about that experiment on the Indiana dunes."

*"It could be May or early June," Enrico told me, "I remember I talked about that experiment on the Indiana dunes."

23. "After a swim in the lake I walked along the beach with Professor Stearns and we talked about the experiment."

*"After a swim in the lake we walked along the beach and I talked about the experiment with Professor Stearns."

24. "We walked ahead on the beach, the two of us, to speak in such a way that the others would not understand."

*"The two of us walked ahead on the beach to speak in such a way that the others would not understand."

25. "Why? Didn't everyone know that you were building piles at the Met Lab?"

*"Why? Didn't everyone at the Met Lab know that you were building piles?"

26. "They knew we were building piles, but they did not know that we had the certainty at last that a pile would work and a chain reaction was feasible.

*"They knew we built piles, but they did not know that at last we had the certainty that a pile would work and a chain reaction was feasible."

Table 1

Mean (and standard deviation) percent ideas in written recall.

Text length and processing	Retrieval interval	
	Same session	1-week delay
13-sentence		
feedback	35.8 (17.7)	19.0 (3.5)
no-feedback	35.2 (11.8)	22.3 (6.1)
26-sentence		
feedback	20.8 (9.5)	16.0 (10.8)
no-feedback	25.0 (8.5)	21.0 (11.2)

Table 2

Mean (and standard deviation) percent sentences in recognition.

Text length and processing	Retrieval interval	
	Same session	1-week delay
13-sentence		
feedback	81.0 (10.4)	86.0 (5.6)
no-feedback	75.7 (13.2)	61.7 (8.4)
26-sentence		
feedback	78.3 (5.5)	59.2 (10.0)
no-feedback	66.7 (14.3)	52.7 (9.4)

Table 3

Mean coefficients of sequencing agreement.

Text length and processing	<u>Coefficient</u>			
	Kendall's tau		Kendall's <u>w</u>	
	Subjects with original		among subjects	
	Immediate	Delay	Immediate	Delay
13-sentence				
feedback	.419	.671	.386	.604
no-feedback	.329	.499	.279	.357
26-sentence				
feedback	.492	.619	.458	.502
no-feedback	.349	.174	.330	.241