

Notetaking and Depth of Processing

BURKE H. BRETZING AND RAYMOND W. KULHAVY

Arizona State University

Four levels of notetaking (summary, paraphrase, verbatim, and letter search) were used to control depth of processing of a prose passage with 180 high school students, who then either reviewed their notes or read an interpolated text. A separate control group took no notes. On immediate and delayed post-tests, post hoc analyses with the depth (notetaking) condition showed the following ranking: summary = paraphrase > control = verbatim > letter search. A paraphrase notes \times review \times test-position interaction was significant, indicating that less forgetting occurred on a delayed post-test when students reviewed their paraphrase notes than when they read an interpolated text. Analysis of reading times showed that the additional time required for notetaking was only worthwhile when meaningful notes were taken.

Written prose is a major source of learning in the classroom, and procedures are needed which facilitate its analysis since even the most conscientious readers may overlook crucial points if they are not paying close attention to the material. Notetaking is a process which can be used to increase a learner's analysis of prose.

When dealing with prose material, perceived meaningfulness is essential for efficient recall. One needs to relate passage content to prior knowledge, in order to facilitate remembering (Johnson, 1973). If students read a passage on professional football, they may evoke images of personal participation or games watched on television. The extent of these images can determine the meaningfulness of the passage to individuals. If the reader is from a country where football is not played, imagery will be sparse, and the passage will not be perceived as meaningful. Thus, a context must be established in memory in order for meaningful learning to take place (Bransford & Johnson, 1972). The present study used a passage about a fictitious African tribe so the specific material would be unfamiliar to subjects; however, common topics were included to provide the material with potential meaningfulness.

Even when prose is potentially meaningful to students, there is no guarantee that they will process the material in a way which facilitates retention. Anderson (1972) notes that semantic encoding is necessary for comprehension, and the learner needs to make a "meaningful representation" of the material to be remembered. Studies have shown that meaningful, semantically encoded material is what a subject recalls after delay (cf. Sachs, 1967). The underlying factor, here, seems to be the process

Requests for reprints should be sent to: Burke H. Bretzing, Department of Educational Psychology, 322 Payne Hall, Arizona State University, Tempe, AZ 85281.

used to ensure semantic encoding. Craik and Lockhart (1972) contend that the process involved in meaningful encoding is attention to content, where closer analysis of text results in more semantic encoding and superior recall over time. Craik and Lockhart coined the title "depth of processing" to describe this model. As people attend to, and analyze a passage more thoroughly, they relate the text to prior knowledge, and encode the material semantically. The greater the depth at which the material is processed, the longer the memory trace.

Several studies have supported the depth-of-processing model (Bobrow & Bower, 1969; Mistler-Lachman, 1974; Schallert, 1976). In one study, Schallert (1976) requested that subjects perform different depth-related tasks when reading a passage of meaningful prose. From most shallow to deepest, these tasks were (a) counting the number of four-letter words in the passage, (b) counting the number of personal pronouns, (c) rating passages for degree of ambiguity, and (d) careful reading. As predicted by the model, greater depth of processing resulted in superior recall. Based on such data, it was hypothesized that subjects in the present study who processed the passage at a deeper level would perform better on both an immediate and a delayed post-test.

A problem arises regarding control of processing depth. How does the teacher deal with the student who simply skims the material and makes no attempt to encode it semantically? Techniques must be provided which assist the student in analyzing the text, and provide concrete evidence that this analysis has been completed. One such procedure is notetaking.

Notetaking is a common procedure in the classroom; especially for recording lectures. After a speech, notes provide the listener with an external storage device, or a source for review of what has been said. Research shows that use of notes for review just before a test can aid recall (Fisher & Harris, 1973; Greene, 1934). Reviewing one's notes can also facilitate delayed retention. Fisher and Harris (1973) showed that reviewers attained higher scores than nonreviewers on a test which followed review by 3 weeks. These results were not significant, but in the right direction.

In terms of depth-of-processing, we are essentially concerned with what the person does during presentation of the stimulus; how the material is encoded. Dyer (1974) found that despite the fact that learners did not review their notes before a 1-week delay test, 84% of the items in their notes were recalled. In this case, subjects probably used notes to encode text, since most of the recalled material came directly from the notes. Therefore, it seems reasonable to assume that notetaking can function as an aid to depth of processing. Since notes are useful for semantic encoding, and deeper processing implies semantic encoding, taking notes should assist in deeper analysis and a longer memory trace.

Notes can be seen as a reflection of how deeply material is being pro-

cessed. If students are reading carefully and relating material to past learning, their notes will probably depict this analytic process. On the other hand, if readers simply skim a passage, their notes may contain only verbatim sentences. It stands to reason, that if notes reflect depth of processing, they should contribute to depth of analysis. If a teacher requests that students take verbatim notes which do not require semantic encoding, learners should not be able to remember as much from the passage as when they are requested to take paraphrase or summary notes which require semantic encoding. The present study provides a test of these relations.

Four levels of notetaking to control depth-of-processing were used in this experiment. From most shallow to deepest, they were: (a) letter-search, recording in notes all words which began with capital letters, (b) verbatim, extracting lines from the passage which were considered to be most critical in the logical structure of the passage to record word-for-word as notes, (c) paraphrase, writing down main ideas from each page as it was read, and (d) summary, recording the main points from each page after it was read. It was expected that notetaking which required deeper processing would result in superior recall, as measured on post-tests consisting of items which required a combination of facts from the passage, and not simple, verbatim answers (Anderson, 1972).

On the basis of prior studies, it was predicted that notes would result in superior recall over no notes, but only when they required semantic encoding (Dyer, 1974; Kulhavy, Dyer, & Silver, 1975). Nonsemantic notetaking, such as letter-searching, would reduce recall, simply because the process requires little attention. These effects would be even more pronounced on a delay measure because meaningful material is best remembered over time (Sachs, 1967).

Of those who did take notes, half reviewed their notes before the immediate test, and half read an interpolated passage during this study interval. Based on previous research, it was hypothesized that reviewers would recall more than nonreviewers, on the immediate test, since the passage was removed after reading, and only reviewers studied their notes (Fisher & Harris, 1973; Greene, 1934). Review was also expected to aid recall on the delayed test (Fisher & Harris, 1973).

Finally, most notetaking studies have simply compared notes with no notes but have failed to investigate differences between types of notetaking strategies. The present study was designed to contrast different notetaking techniques, and suggest methods which might be used to facilitate learning from them.

METHOD

Design. Two factors, Depth-of-Processing and Review, were combined factorially to form eight experimental cells. Test-Position was treated as a within subjects factor across each experimental cell. Hence, the design was a 4 Depth-of-Processing (summary vs para-

phrase vs verbatim vs letter-search) \times 2 Review (review vs no-review) \times 2 Test-Position (immediate vs delay) design with repeated measures on the Test-Position factor. A separate read-only group also completed the text and took the tests.

Subjects. A total of 180 male and female junior and senior student volunteers from a local high school served as subjects. Twenty students were randomly assigned to each group in order of their appearance for the experiment. Eighteen subjects were eliminated due to failure to complete all tasks, leaving 18 per cell.

Materials. The text materials consisted of a 2000-word passage titled *The Gruanda* which describes an imaginary African tribe (Anderson & Myrow, 1971). The text covers topics in the geography, organization, physical characteristics, religion, rituals, economics, and government of the tribe. The passage was typed double space on eight 8.5 \times 11 (22 \times 28 cm) pages of white foolscap. Anderson and Myrow (1971) previously reported an uncorrected readability level of 6.6 for the passage, using the Farr, Jenkins, and Patterson revision of the Flesch Reading Ease Formula (Klare, 1963). Sheets of 8.5 \times 11 (22 \times 28 cm) white foolscap with three lines per page were used for notetaking.

A passage on introductory psychology was used as the interpolated task. This passage contained no information covered in the experimental text (Anderson & Myrow, 1971).

A 25-item constructed-response criterion test was developed to test for knowledge of information obtained from the passage. The answers to these questions were evenly distributed throughout the passage, and items were randomly arranged so they would not be in the same order as they appeared in the text. The questions were of a substantive, integrative nature, requiring a combination of facts from the passage, and not a simple, verbatim answer. The questions could be answered in a short sentence. The items are similar to those used by Anderson and Myrow (1971).

Correct answers for the items were taken directly from the passage. One point was awarded for correct answers, and no points if the answers were incorrect. Errors in spelling and grammar were allowed, but to receive credit for an answer, students had to use the correct substantive words from the passage. All tests were scored by the same person.

Procedure. The experiment was conducted during two different sessions. The first session included reading, notetaking, and review, followed by a test, and the second simply repeated the testing procedure.

Groups of subjects ranged in size from 20–40 with all experimental conditions represented at each session. As subjects entered the classroom, they were given booklets appropriate to their experimental task, consisting of instructions, the passage, and the lined paper. One sheet of lined paper was used with each page of the text. Subjects were told to sit down and read the cover sheet instructions carefully. Specific instructions were presented on the second page of each booklet. Verbal instructions told learners not to be concerned about what other students were doing, because there were several different tasks randomly located about the room. Subjects were also told that a test would follow. After questions were answered, readers were told that they would be given 30 min to read the passage, and they were to raise their hands if they finished before this maximum time. A norming study suggested 30 min as the maximum time required to read and take notes on the passage. Times were written on the board at 15-sec intervals.

Written instructions told all subjects to read the passage carefully. Notetakers were encouraged to use all three lines on the separate pages for notes from each page of the text, but to limit themselves to only three lines. When all subjects had finished reading their specific instructions and asking general procedural questions, they were told to begin reading the passage.

Subjects in the summary group read each page of the text and wrote three lines on the attached lined paper that they felt was a good summary of the main ideas of the entire page.

Students in the paraphrase group took notes as they read. These subjects took notes which they felt expressed the main ideas of the passage and were most important to the logical structure of the text. As subjects in this group read each sentence, they decided if it met the above criteria, and if it did, they put it in their own words and wrote their paraphrase on the lined paper.

Learners in the verbatim group read each page of the text and selected from it three lines which they felt were the most critical in the logical structure of the page. As subjects read through the text of each page, they picked exactly three lines which they thought were the most essential and wrote them on the attached lined paper.

The letter-search group located all words on each page which began with capital letters, and wrote them on the attached sheets. These subjects were told to fill the entire three lines with words, even if it was necessary to repeat some.

Finally, the control group was given written instruction to simply read the passage carefully and raise their hands when they finished.

As readers raised their hands, a monitor went to them and recorded their reading times. If subjects were in the control group, their passages were collected and they were given the interpolated passage to read for 5 min. If subjects were in the no-review group, both their passages and their notes were collected, and they were given the interpolated passage to read for 5 min. If the subjects were in the review group, their passages were collected, and written instructions told them that they would be given 5 min to study their notes in preparation for a test which would follow. All subjects were told that a monitor would tell them when their 5-min study interval had elapsed.

As students completed their study interval, either their notes or their interpolated passages were collected, and they were given the post-test. Written instructions told subjects that they would be given as much time as they needed to complete the test. It was requested that every item be answered, even if it was only a guess. When the exams were completed, the papers were collected and the subjects allowed to leave. The only exception to this procedure was that learners in the control group finished much sooner than others, since they did not take notes. So, following the post-test the interpolated passage was returned to these subjects. This group read until notetakers began to finish the post-test, and then they were allowed to leave.

After 1 week, the subjects returned and were given the delayed test. The test items were identical to those on the immediate test, but were separately randomized. As before, students were asked to answer all items to the best of their ability, and there was no time limit.

RESULTS

The immediate and delay post-tests were scored according to the criteria established above. One point was awarded for correct answers, no points if the answers were incorrect, and number right out of 25 was recorded for each subject. Reading times, in minutes, for the experimental passage were also obtained.

Table 1 depicts the individual cell means and standard deviations for number correct on both the immediate and delay post-tests. A 4 (Depth-of-Processing) \times 2 (Review) \times 2 (Test-Position) analysis of variance with repeated measures on test-position produced significance for the depth-of-processing, $F(3,136) = 8.16, p < .01$, and test-position, $F(1,136) = 335.53, p < .01$, main effects and for the depth-of-processing \times review \times test-position interaction $F(3,136) = 3.76, p < .05$.

TABLE 1
MEANS AND STANDARD DEVIATIONS FOR BOTH IMMEDIATE AND DELAY CORRECT RECALLS

Review	Depth-level	Test-position			
		Immediate		Delay	
		\bar{X}	SD	\bar{X}	SD
Review	Summary	13.56	4.14	10.00	4.08
	Paraphrase	12.83	4.91	11.00	5.50
	Verbatim	10.22	3.33	6.94	4.07
	Letter search	9.00	4.59	5.94	4.49
No review	Summary	13.94	5.27	10.55	5.04
	Paraphrase	14.11	4.04	9.94	4.63
	Verbatim	11.11	4.95	8.33	4.67
	Letter search	9.44	4.44	6.33	3.10
	Control	11.11	6.00	8.17	5.28

Post hoc Newman-Keuls comparisons were performed on the depth-of-processing means, resulting in the following ranking of conditions: summary = paraphrase > verbatim > letter-search ($p < .01$). Apparently, it made little difference whether students took summary or paraphrase notes, but both of these semantic levels of processing were superior to word-for-word notes, which were, in turn, more valuable in recall than simply writing down capitalized words.

Comparisons between the control group and the depth-of-processing conditions were made using Dunnett's tests. The control group's scores were significantly lower than both the summary and paraphrase group's ($p < .01$), not significantly different from the verbatim subjects' scores, but significantly higher than those of the letter-search group ($p < .05$). Thus, people who took summary and verbatim notes recalled more overall than students who took no notes, who were about the same as verbatim notetakers, and verbatim and no notes groups were better in recall than letter-search notetakers. Combining the Newman-Keuls and Dunnett's comparisons, the following ranking emerges: summary = paraphrase > control = verbatim > letter-search.

In order to determine the source of the significance within the review \times depth-of-processing \times test-position interaction, a simple effects analysis was completed. Contrasts were made at each level of every factor, to discover which particular interactions were being combined to form the overall interaction. The results demonstrate a paraphrase notes \times review \times test-position interaction to be the only one to reach significance, $F(1,136) = 6.41$, $p < .01$. Therefore, when semantic (paraphrase) notes were taken, it made no difference on the immediate post-test whether students reviewed their notes or read an interpolated passage, but there

was significantly less forgetting after a 1-week delay when students reviewed their notes before the first test.

For reading times, a Dunnett's test was used to compare the mean of the control group with those of the remaining conditions. All comparisons were significant ($p < .01$), so notetakers took significantly longer to read the passage than nonnotetakers.

DISCUSSION

In this study, students who took notes which required summarizing or paraphrasing recalled significantly more than those who took either verbatim or letter-search notes. Both summary and paraphrase activities were designed to control semantic levels of processing, because they require learners to integrate new material with prior knowledge (Johnson, 1973). However, the meaningful encoding process does not seem to be a factor in the lower levels of notetaking.

No significant difference was found between summary and paraphrase notetaking. In each case, subjects were requested to select what they felt were the most important points from each page and record them as notes. Thus, both levels of notetaking can be combined and called semantic processing, as they both were used to control meaningful encoding with similar results.

Semantic processing can then be compared with verbatim notetaking. Since no meaningful encoding was required to complete the verbatim task, it was considered to be a nonsemantic level of processing. The results on both immediate and delay tests confirmed these expectations. Subjects who took verbatim notes scored significantly lower on tests of comprehension than those processors at a higher level of depth. Clearly, the verbatim group was not attending to the content of the passage as closely as the summary and paraphrase groups.

The lowest level of processing occurred in the letter-search condition. These notetakers produced significantly fewer post-test corrects than either the verbatim or semantic groups. Letter-search subjects seemed to take the easiest route and simply wrote words without paying much attention to what they were reading.

The scores of the control subjects were about the same as verbatim notetakers on both immediate and delayed post-tests. Two conclusions can be drawn. First, writing sentences word-for-word did not aid recall any more than reading alone, and both were superior to writing capitalized words. Second, semantic notetaking increased recall over just reading. Therefore, when reading a passage, verbatim notetaking does not increase performance on comprehension tests, but summary and paraphrase notetaking do. As expected, letter-searching is a level of analysis which actually detracts from reading alone in terms of test performance.

The review main effect was not statistically significant. It made little difference whether subjects reviewed their notes or read an interpolated passage during a 5-min interval before the immediate post-test, and except in one case which will be discussed below, there was little difference on the delayed post-test. One explanation for the lack of effects on the immediate post-test is due to the simplicity of the passage. The text may have been so clear to readers, especially in the semantic processing conditions, that during a 5-min delay, there was no advantage to reviewing notes. The interpolated passage was noninterfering, so for semantic notetakers who clearly understood the passage, a short delay made no difference.

The paraphrase notes \times review \times test-position interaction was statistically significant. On the immediate post-test, for the paraphrase condition, there was no appreciable difference between reviewers and non-reviewers on total corrects. Both groups had just thoroughly analyzed the passage and additional study was probably unnecessary, but during the 1-week delay, the paraphrase reviewers forgot significantly less material than the nonreviewers. Semantic notetaking facilitates associations with prior knowledge, which is strengthened when coupled with review. Fisher and Harris (1973) found that notetakers who reviewed their notes after the passage, but not before the delayed test were superior to those who took notes but did not review. These results were in the right direction, but not significant. The present study provides significant data for the benefits of review.

Students who did not take notes required significantly less time to complete the text. Craik and Lockhart (1972) state that a longer period of time spent with material will result in superior recall if this time is used for a deeper analysis. Their hypothesis is supported by this study. All notetakers spent significantly more time reading the passage than the controls, but only summary and paraphrase notetakers had significantly higher recall on the post-tests.

In conclusion, this study has demonstrated that notetaking can facilitate recall of prose. Taking notes requires more time than reading alone. The greater amount of time needed for notetaking becomes worthwhile only if the notes are of a semantic nature which increases the depth at which material is processed, thereby strengthening the memory trace.

REFERENCES

- ANDERSON, R. C. How to construct achievement tests to assess comprehension. *Review of Educational Research*, 1972, 42, 145-170.
- ANDERSON, R. C., & MYROW, D. Retroactive inhibition of meaningful discourse. *Journal of Educational Psychology Monograph Supplement*, 1971, 62, 81-94.
- BOBROW, S. A., & BOWER, G. H. Comprehension and the recall of sentences. *Journal of Experimental Psychology*, 1969, 80, 455-461.

- BRANSFORD, J. D., & JOHNSON, M. K. Contextual prerequisites for understanding: Some investigations of comprehension and recall. *Journal of Verbal Learning and Verbal Behavior*, 1972, 11, 717-726.
- CRAIK, F. I., & LOCKHART, R. S. Levels of processing: A framework for memory research. *Journal of Verbal Learning and Verbal Behavior*, 1972, 11, 671-684.
- DYER, J. *Effects of note-taking and repetition on encoding and retention of printed matter*. Unpublished doctoral dissertation, Arizona State University, 1974.
- FISHER, J. L., & HARRIS, M. B. Effects of note-taking and review on recall. *Journal of Educational Psychology*, 1973, 65, 321-325.
- GREENE, E. B. Certain aspects of lecture, reading, and guided reading. *School and Society*, 1934, 39, 619-624.
- JOHNSON, R. E. Meaningfulness and the recall of textual prose. *American Educational Research Journal*, 1973, 10, 49-58.
- KLARE, G. C. *The measurement of readability*. Ames: Iowa University Press, 1963.
- KULHAVY, R. W., DYER, J. W., & SILVER, L. The effects of note-taking and test expectancy on the learning of text material. *Journal of Educational Research*, 1975, 68, 363-365.
- MISTLER-LACHMAN, J. Depth of comprehension and sentence memory. *Journal of Verbal Learning and Verbal Behavior*, 1974, 13, 98-106.
- SACHS, J. S. Recognition memory for syntactic and semantic aspects of connected discourse. *Perception and Psychophysics*, 1967, 2, 437-442.
- SCHALLERT, D. L. Improving memory for prose: The relationship between depth of processing and context. *Journal of Verbal Learning and Verbal Behavior*, 1976, 15, 621-632.