PRODUCTION NOTE

University of Illinois at Urbana-Champaign Library
Technical Report No. 153

COMPREHENSION MONITORING
AND THE ERROR DETECTION PARADIGM

Peter Winograd and Peter Johnston
University of Illinois at Urbana-Champaign

January 1980

Center for the Study of Reading

UNIVERSITY OF ILLINOIS
AT URBANA-CHAMPAIGN
51 Gerty Drive
Champaign, Illinois 61820

BOLT BERANEK AND NEWMAN INC.
50 Moulton Street
Cambridge, Massachusetts 02138
Technical Report No. 153

COMPREHENSION MONITORING
AND THE ERROR DETECTION PARADIGM

Peter Winograd and Peter Johnston
University of Illinois at Urbana-Champaign

January 1980

University of Illinois
at Urbana-Champaign
51 Gerty Drive
Champaign, Illinois 61820

Bolt Beranek and Newman Inc.
50 Moulton Street
Cambridge, Massachusetts 02138

The research reported herein was supported in part by the National Institute of Education under Contract No. US-NIE-C-400-76-0116.
This page is intentionally blank.
Abstract

The purpose of this study was to examine the conditions which were likely to facilitate error detection. It was hypothesized that poor readers' comprehension monitoring abilities would improve if they were given assistance in selecting the appropriate schema for understanding a passage. In order to test the hypothesis, we used a standard paradigm: the error detection task. No evidence was found to support the notion that schema activation would significantly improve poor readers' error detection abilities.

However, results did indicate that, while good readers were significantly better at this task than were poor readers, a surprising number of children failed to report some very blatant errors. Although these results are in agreement with earlier studies using the same task, we felt uneasy in drawing the conclusion that sixth graders are lacking in metacognitive abilities.

Instead we have expanded the discussion to include our thoughts on the limitations and difficulties in the use of the error detection paradigm itself. Five major concerns were identified and suggestions for improving future comprehension monitoring studies were made. Some alternative methodologies were also considered.
Comprehension Monitoring

There is a great deal of current interest in metacognition. This interest is based on the realization that a valuable distinction can be made between knowledge (cognition) and the awareness and control of that knowledge (metacognition). Vygotsky (1962), for example, described two phases in the development of knowledge: first, the automatic, unconscious acquisition of knowledge; and second, the gradual increase in the active control over that knowledge. Brown (1978) defended the isolation of metacognitive skills for study on the grounds that such skills are the essence of intelligent activity. Research is underway in many subareas of metacognition. Reviews of the relevant literature can be found for: (a) metacognition and memory (Brown, 1978; Flavell & Wellman, 1977); (b) metacognition and linguistics (Clark, 1978); (c) metacognition and reading (Anderson, in press; Brown, in press; Markman, Note 1); (d) metacognition and communication (Shatz, 1978); (e) the social origins of metacognition (Wertsch, Note 2).

This paper is concerned with metacognition and reading. More specifically, it is concerned with comprehension monitoring and one of the research paradigms currently in vogue, error detection. It is our contention, gained through hindsight, that there are some serious problems which may limit the usefulness of any data that are collected by this method. The first part of this paper will review the original thoughts that led to our attempt to use the error detection paradigm. Then the study itself will be described. The last section will cover the issues and difficulties that forced us to reconsider methodologies.
Metacognition and Reading

Metacognition has been defined as "knowledge that takes as its object or regulates any aspect of any cognitive endeavor" (Flavell, Note 3). Some researchers interested in the relationship between metacognition and reading have focused on what readers know about the task of reading, while others have looked at how readers regulate and monitor ongoing processes during reading. These two lines of research reflect a difference in emphasis, not two independent entities (Brown, Note 4). Accordingly, some researchers have attempted to study both aspects of metacognition.

The study by Canney and Winograd (1979) is an example of the first kind of research. The main purpose of that study was to see if a reader's perception of the purpose of reading (decoding vs. meaning getting) was related to reading ability. Interview results indicated that many poor readers may share the same perspective on reading as the fourth-grader who said that good reading is being able to say all the words fast. While these results are tentative, they do indicate that there may be an important relationship between a child's view of reading and his ability to perform. Other studies in this vein have looked at such things as children's ability to explicitly identify important aspects of text (e.g., Brown & Smiley, 1977; Otto, Barrett, & Koenke, 1969; Stein & Glenn, 1978), or the child's view of what constitutes a word (Downing & Oliver, 1973-74).

The focus of this paper, however, is on the research that deals with how readers regulate and monitor ongoing processes during reading. The question under consideration concerns the conditions under which readers monitor how
well or how poorly their comprehension is proceeding. This kind of comprehension monitoring is important because it provides vital feedback to the reader about the effectiveness of his reading behavior. A reader who is monitoring his own comprehension has a better basis for selecting the reading strategy best suited to the needs of the moment. Indeed, it might be argued that the definition of a fluent reader must include a reference to the ability to self-check and self-correct reading strategies.

In order to understand the factors involved in comprehension monitoring, some researchers have used error detection tasks. Such tasks usually involve reading (or listening to) a passage in which an error has been embedded. If the subject does not mention the error following the reading, probe questions are often asked in an attempt to learn why. Error types have included disorganized passages (Danner, 1976), incomplete instructions (Markman, 1977), inappropriate transition words linking sentences and unclear pronominal references (Baker, 1979), and contradictory information (Baker, 1979; Markman, 1979). Subjects have included both children and adults. Although the types of errors have been different for adults and children, neither age group has performed well in any of their respective error detection tasks. This is an important point, and we will return to it later.

There are many possible explanations of the factors involved in comprehension monitoring. Not surprisingly, these explanations often reflect the theoretical biases of the researchers offering them. The remainder of this section describes a schema-theoretic perspective of comprehension monitoring, and the next section describes an experimental study aimed at
addressing the issues of schema activation and comprehension monitoring. The final section of the paper will examine, retrospectively, some of the problems in the error detection paradigm and our reasons for selecting additional methodologies for future studies.

The Role of Schemata in Comprehension Monitoring

It is important for theoretical, as well as practical, reasons that we understand the cognitive mechanisms that are involved in comprehension monitoring. A basic assumption in current information processing theory is that a human's processing capacity is limited and that cognitive processes must compete for space (Norman & Bobrow, 1975). Therefore we develop ways which allow us to use our limited processing capacity more efficiently. One way is routinization. Cognitive processes that are routinized are able to be processed with a minimum of effort, freeing valuable processing capacity.

Since reading comprehension is a cognitive process, it too is involved in the race for space. In some situations comprehension is routinized. "Comprehension in the normal case is a fully automatic process, that is, it makes low demands on resources" (Kintsch & van Dijk, 1978, p. 372). But reading comprehension should be considered as a continuum with routinization at one end. At the other end, comprehension can be a slow laborious process which takes our undivided attention. Comprehension monitoring allows us to determine whether our position on the continuum is an effective one. One way in which this is accomplished is by means of a subjective feeling that is experienced as "confusion, uncertainty or some similar sensation which informs us that we have failed to understand" (Markman, 1979, p. 1).
In order to understand how comprehension monitoring works, it is helpful to understand what happens when comprehension is proceeding smoothly. Rumelhart (1977) has developed a definition of comprehension that is most useful here. Comprehension is the "process of selecting and verifying conceptual schemata to account for the situation (or text) to be understood" (p. 268). A simplified explanation is that schemata provide a general outline and the reader fills in the specific details from the text.

Schema theory provides a framework for examining how a reader becomes aware of a comprehension failure. Consider the possibility that error detection results when there is enough of a mismatch between input and the selected schemata to induce the subjective feelings of confusion. In other words, properly selected schemata provide the reader with certain expectations. When the incoming information fails to conform to these expectations, the monitoring processes signal trouble.

Recall the point raised earlier, that one of the consistent findings of the comprehension monitoring studies is that many readers fail to mention the errors embedded in the text. One possible explanation is that some readers may not have selected the appropriate schemata. Poor readers, in particular, may suffer from this problem. There is a considerable body of evidence that poor readers often use highly inappropriate schemata. In addition to the Canney and Winograd (1979) study mentioned earlier, the interested reader should consult Downing (1969), Glass (1968), Johns (1974), Johns and Ellis (1976), or Weintraub and Denny (1965).
The following study was designed to explore the possibility that good readers are more likely to select appropriate schemata, which help in comprehension monitoring, than are poor readers. One way to test this possibility is to see whether poor readers' error detection abilities improve when they are given assistance in selecting appropriate schemata.

Method

Subjects

Twenty-sixth grade students were divided into skilled and less skilled readers on the basis of their scores on the Stanford Diagnostic Reading Test, Form A. Skilled readers were defined as those who scored above the 50th percentile nationally. Their scores had a mean of 80, SD of 16, and ranged from 53% to 96% (n = 9). Less skilled readers were defined as those who scored below the 50th percentile nationally. Their scores had a mean of 31.1, SD of 11.13, and ranged from 16% to 46% (n = 11).

As an assurance that all the subjects could decode the words in the passages, each subject read a word list which contained, in random order, all the words involved in the experimental paragraphs. Each word appeared only once in the list. Accuracy in this decoding task was sufficiently high to utilize the data of all subjects in the analyses.

Materials

The materials consisted of four ten-sentence paragraphs, two dealing with a circus theme and two dealing with a church theme (see Appendix). All were similar to Bransford and Johnson's (1973) Peace March/Space Travel paragraph
in that they were intentionally ambiguous. A single reference to either circus or church occurred in the sixth sentence.

The error type used in this experiment was a contextually anomalous sentence. This anomalous sentence occurred as sentence number 8 in each paragraph.

Design

There were two major factors under consideration. The first factor (between-subjects) was reading ability. The second factor (within-subjects) was the degree of contextual preparation supplied prior to the reading of the paragraphs. There were two levels of the second factor:

1. No Preparation
2. Preparation

The preparation was a production task in which the subject looked at a picture of two children approaching either a church or a circus (see Appendix) and was then asked to "Use your imagination and tell me everything you think these two children might see in this circus/church."

During the no-preparation session the children were asked to read the previously described word list.

Procedure

The children were tested individually in each of the two sessions, given a week apart. The sessions were tape recorded so the children did not have to write at any time. They were asked to act as consultants in determining the comprehensibility of some passages written by other sixth graders. They then
received either the preparation or the no-preparation task. Next, they were
told to silently read the first of the paragraphs through twice. When they
had finished reading, they were asked a series of probe questions to see if
they had detected the target anomaly. The children then read the second
paragraph (same theme) which was again followed by the probes.

During the second session, the children were again asked to act as
consultants, and the procedure was repeated with the two remaining paragraphs.
The order of the preparation/no-preparation task and the theme of the
paragraphs were counter-balanced across both sessions, while the paragraphs
themselves were counter-balanced within each session.

The children's reactions to the paragraphs were monitored closely. If a
child made any comment indicating he had detected the target anomaly during
the initial reading of the passage, he was assigned a score of 1. Otherwise,
a child's score represented the number of the probe that did produce a
relevant comment. As soon as a child detected the error, the probes were
discontinued. However, all the children were asked probe #11, "Tell me what
the story was about." If a child failed to comment on the target anomaly
throughout the probes, he was assigned a score of 12.

The probes were:

2. Any comments?
3. Do you have any questions?
4. What did you think about the story?
5. Did everything make sense?
6. Could fifth graders understand everything?
7. Could you answer questions about it?
8. Would you change anything in the story?
9. Does everything sound all right?
10. Are there any sentences that don't fit?
11. Tell me what the story was about.

In summary, then, several possible factors underlying comprehension monitoring were considered. Two of these factors were of major interest: (a) reading ability; (b) assistance in selecting the appropriate schemata (preparation) or no assistance (no-preparation). In addition to these major factors, story, story order, and session order were examined for their possible effects. The dependent measure was the child's score. This variable could range from 1 to 12 depending on whether the child mentioned the embedded error during the initial reading of the passage (1), or in response to one of the probes (2-11), or failed to mention the error at all (12).

Results

The initial data analysis was a 2 x 2 x 2 x 2 repeated measures analysis of variance. The between-subjects factors were reading ability (high vs. low), treatment order (preparation first or no-preparation first), and story order within each session. Preparation vs. no-preparation (the experimental treatment) was measured within subjects. Table 1 summarizes the two ability and the two treatment conditions.

The analysis indicated that good readers performed significantly better than did the poor readers, $F(1,16) = 4.89, p < .05$, but that, contrary to
prediction, schema preparation prior to reading had no significant effect on error detection. The only other significant effect in the initial analysis was the interaction between treatment and the order in which the subjects received the treatment (preparation first or no-preparation first), $F(1, 16) = 7.42, p < .05$. Figure 1 summarizes this interaction.

![Insert Figure 1 about here](image1)

One explanation for this interaction is that the preparation first group was comprised of better readers than the no-preparation first group and thus did consistently better over the two sessions. However, a $t$-test which compared the reading comprehension scores of the two groups indicated that the two groups did not significantly differ from one another, $t(18) = 1.19, p > .05$.

![An additional analysis was performed to clarify the order-of-presentation effects. Since there was no main effect for treatment, a $2 \times 2 \times 2$ repeated measures analysis of variance was run on the factors of reading ability, session (first or second), and story order within session. Figure 2 clearly shows the significant effect of session, $F(1, 36) = 6.3, p < .05$, as well as the significant effect of reading ability, $F(1, 36) = 6.05, p < .05$.](image2)

Simply put, good readers did better than poor readers, and both groups did better the second time around.
One potential problem in the initial analysis was the excessive error variance associated with using the twelve probes as the dependent measure. Many subjects received the highest score (1) on the first paragraph and then the lowest score (12) on the next paragraph, or vice versa, within the same session. For this reason, we thought that a new scaling system was needed which would reduce the variance and perhaps reveal other patterns in the data. A new scaling system was devised which awarded a 1 if the subject spontaneously mentioned the target anomaly; a 2 if any of the probes resulted in an error detection response; and a 3 if the subject failed to mention the target error at all. The original $2 \times 2 \times 2 \times 2$ repeated measures analysis of variance was rerun using this scaling system. Reading ability approached significance, $F(1,16) = 3.94, p < .06$, and the significant treatment-by-order interaction increased to $F(1,16) = 8.71, p < .01$. The new results differed little from the results of the first analysis, indicating that the variance had not obscured other possible effects.

In order to examine the materials used in the experiment, the effects of the treatment were considered for each story. The cell means are shown in Table 2.

An analysis of variance (Winer, 1962, pp. 635-639) indicated that, while there was still no main effect for treatment, there was a significant main effect for stories, $F(3,49) = 3.98, p < .05$. Post-hoc analysis (Scheffe) indicated that story 3 was significantly different from stories 1, 2, and 4, $p < .05$. 

---

Insert Table 2 about here

---
and that stories 1, 2, and 4 did not differ significantly from one another. Tables 3, 4, 5, and 6 represent the ability by treatment conditions for each story separately. Performance on stories 1 and 4 was as predicted, whereas on story 3 results were in the opposite direction. Story 2 was inconclusive.

The efficiency of the probes used as the measurement instrument was examined via the distribution of frequencies of triggering a correct error detection response. Table 7 shows this information.

It appears that probes 6, 7, and 9 failed to elicit any error detection responses at all, and that probes 3 and 8 functioned minimally. Probes 2, 5, and 10 seemed to function best. Seventy-three percent of the total responses were made either during the initial reading of the passage (41%) or during the first four probe questions (31%). Thus, the early probes seemed to be more effective than did the later ones. It is interesting to note that probe #10, "Are there any sentences that don't fit?" accounted for 23% (5/22) of the possible remaining responses.

Discussion

This discussion is divided into three parts. First, the results of the schema activation experiment will be discussed. Second, we will focus on the difficulties and limitations in the use of the error detection paradigm itself. The third part will be concerned with some of the alternative
paradigms which are currently in use, and with some suggestions for improving the methodologies used in future comprehension monitoring research.

A surprising number of the sixth graders were quite poor at overtly detecting some very blatant errors. Only 56% of the good readers and 18% of the poor readers mentioned all four errors. If story 3 is excluded from this analysis, 88% of the good readers and 45% of the poor readers mentioned all three of the target anomalies. Good readers, as expected, did significantly better than did poor readers, but neither group did outstandingly well on the error detection task. While these results support those of earlier studies (Baker, 1979; Markman, 1977, 1979), they leave unanswered a central question: Did the children fail to detect the errors or did they just fail to mention them? We attempted to increase communication by having a period of discussion with each child to establish rapport, and by telling the children that the stories were written by other sixth graders. There was evidence in the protocols that most of the children criticized some aspect of the stories. However, the criticism of unintelligibility is a major one and may be perceived as being a reflection on the reader's ability. Thus the question of whether or not the children detected the error is still unanswered. The difficulty of answering this question is one of the major problems of the error detection paradigm and will be discussed in depth later.

The only other statistically significant finding in the initial analysis was the interaction between treatment and the order in which the children received the treatment. As reported earlier, one possible interpretation was that the group of children who received the preparation treatment during the
first session were simply better readers overall than the children who received the no-preparation treatment first. However, the non-significant results of the t-test, which compared the comprehension test scores of the two groups, argues against this interpretation. The interpretation we favor is that those children who received the preparation in the first session benefited from it and that this advantage, combined with the general order-of-session effect, carried over into the second session. Although the treatment effects were not significant, this plausible interpretation of the data does offer support for our original hypothesis.

An additional analysis was performed on the factors of reading ability, order of presentation, and story order within each session. The results from this analysis clearly showed the effects of session order. The children did better in the second session. Whether this improved performance resulted from a better understanding of the task or an increased rapport with the experimenter is hard to determine. Either explanation supports the need for having a longer warm-up session as well as insuring that the children fully understand the task.

Overall, it appears that the treatment had no significant effect. A more fine-grained analysis shows why. Data from two of the stories (1 and 4) provided the expected results. Data from story 2 was inconclusive. Those from story 3 produced some very strong results in the opposite direction. While our hypotheses predicted that all readers would be more likely to spot the errors in the preparation condition and that this would be more so for the poor readers, the results from story 3 indicated quite the reverse. Both
groups did worse in the preparation condition and this was especially true for poor readers (See Table 5). Apparently, some children assumed that the teacher was at Sunday School, thus rendering the intended anomaly quite meaningful. Story 3 emphasizes the need for ensuring that all of the embedded errors are as similar as possible, both qualitatively and quantitatively. This problem will also be discussed further in the next section.

It is interesting to note that the preparation condition enhanced this Sunday School inference for Story 3 for both good and poor readers. However, the effect was stronger for poor readers. These data can be interpreted as tending to confirm the original hypothesis in that the greater the schema preparation, the more easily readers were able to find a "slot" for such ambiguous information. These data also raise another interesting question: Do poor readers' inferencing abilities improve when they are given assistance in selecting the appropriate schema?

Another finding of importance concerned the probes. Several of the probes failed to elicit any error detection responses. Unfortunately, these same probes are some of the most intuitively appealing. It is unclear whether the problem lies in the probes themselves or in the order in which they were given, although the evidence (see Table 7) does suggest that earlier probes are more effective. The probes also caused other problems. Some of the children became impatient with the experimenters when they were asked what was basically the same question several times. It is also very difficult to accurately assess what training effect the probes produced. It is possible that some of the children's purposes for reading were altered as the probes
became more specific. It is also possible that the probes had an influence on the children's perception of the experimental task demands. Evidence for this last point can be found in the fact that probe #10, "Are there any sentences that don't fit?" elicited a large number of error detection responses from children who, up to that point in the interview, had not mentioned the error.

This study failed to find any significant effect of schema activation on children's error detection abilities. However, we feel that the error detection paradigm did not allow for a strong test of the hypothesis. It will prove instructive to look at some of the difficulties and limitations of the error detection paradigm which formed the basis for this conclusion. Several of these problems have already been mentioned. These, along with some others, will now be considered in detail.

The most serious limitation of the error detection paradigm is the difficulty of determining why subjects do so poorly on the task. One cannot say for certain that the subjects' comprehension monitoring abilities are poor because they failed to mention the error. There are numerous other explanations. Some of these have been mentioned in the literature (Baker, 1979; Markman, 1979). For example:

1. A reader's lack of relevant background knowledge may cause him to overlook the error.

2. Readers, especially young ones, may suspend disbelief because they have read much that is unbelievable.

3. Older readers may have an overriding faith in the Cooperative Principle (Grice, 1975), which states that speakers and writers
usually intend their messages to be truthful, relevant, and unambiguous.

4. Readers may not believe that texts can and often do contain errors.

5. Readers may make inferences that seem to resolve the errors and discrepancies.

6. Subjects (young children) may not recall the inconsistent information (Markman, 1979).

7. Subjects may lack the logical capacity to make the necessary inferences (Markman, 1979).

8. Subjects may be hesitant to criticize the experimenter in a testing situation (Markman, 1979).

9. Subjects may draw upon prior knowledge to supplement explicitly presented information (Baker, 1979).

10. Subjects may assign alternative meanings to the text (Baker, 1979).

11. Subjects may assume that the writer has made a mistake and ignore it (Baker, 1979).

12. Subjects may notice the error but assume that subsequent information would resolve the problem (Baker, 1979).

Thus to assume that metacognitive abilities are not well developed because subjects do poorly on the error detection tasks is unwarranted.

Another problem, closely related to the first, involves setting a criterion for deciding when comprehension is adequate. The difficulty arises when the experimenter assumes, either implicitly or explicitly, that the subject's purpose for reading matches his own. Baker (1979), for example,
reports that some of her subjects did not respond to the target error because they were reading for the main ideas. They felt that they had understood the central theme even if one of the sentences seemed to be in conflict.

It is likely that comprehension monitoring is intimately tied to the reader's purpose for reading. Unless some effort is made to accurately assess or control for the subject's purpose for reading, it is difficult to make any generalizations about a reader's ability to detect errors.

The third problem with error detection studies involves specifying the kind and degree of target error. Consider some of the different categories of errors that can be used:

1. Omissions
2. Inconsistencies
3. Unclear pronominal references
4. Inappropriate transition words
5. Disorganized passages
6. Anomalous sentences
7. Spelling errors
8. Grammatical errors

It should also be noted that there is a wide choice of error type within each of these categories.

It may be that there is an error type by subject by task demand interaction which again limits the usefulness of any generalization. Danks and his colleagues (Danks, Fears, Bohn, & Hill, Note 5) offer evidence that some kinds of errors may affect comprehension processes which are text driven.
while other kinds of errors may affect those comprehension processes which are schema driven. If, as Spiro and Tirre (1979) have suggested, readers differ in their use of text-based or knowledge-based information, then the possible interaction between individual differences and error types needs to be considered.

Just as the kind of error is important, so too is the location of the error. Baker (1979) reported that errors placed high in the text structure were noticed more often than the same errors placed lower in the hierarchy.

The fourth limitation with the error detection paradigm lies in the use of probes to see if the subjects noticed the errors. Two of these problems were mentioned earlier. First, some of the children became impatient when they were repeatedly asked for their opinions about the stories. Ten probes may be just too many to ask. The second problem with the probes lies in accurately assessing what effect they are having on the children’s performance. Probes which ask for specific information may induce a different comprehension set than do probes which are more general in nature. A third problem arises when the probes are used as the basis for the scoring procedure. In this study, for example, the scoring procedure assumes equal intervals between probes. Although this procedure has been used before (Markman, 1977, 1979), it is difficult to justify on statistical grounds.

The fifth weakness in the error detection paradigm is that it relies on subjects to make verbal reports about their cognitive processes. Nisbett and Wilson (1977) argue persuasively that there may be little or no direct introspection of higher order cognitive processes. Instead, subjects base
their reports on their judgments of what they consider to be plausible causes rather than on true introspection. This problem is compounded in the error detection paradigm because the probes may "lead" a subject to certain conclusions. Brown (1978) has also illustrated the difficulties in accepting verbal reports from young children. What children (and adults) say is often different from what they actually do.

These, then, are the major difficulties in using the error detection paradigm:

1. Determining why subjects do not overtly respond to the presence of errors in the text.
2. Determining which criteria for comprehension subjects have chosen to apply.
3. Adequately specifying the kind, magnitude, and placement of the target errors.
4. Overrelying on the use of probes as the dependent measure.
5. Overrelying on subjects' (especially children) verbal reports about their own cognitive processes.

The existence of these problems does not mean that the error detection paradigm is useless and should be avoided. However, it does mean that a great deal of thought should be given to which research questions can be addressed by such a methodology. This paradigm is at its weakest when it is used to answer such general questions as: "Do embedded errors affect the reader?" or "How is the reader affected?" It is at its best when it is used to assess a reader's ability to overtly report the effects of embedded errors.
important to note that the error detection paradigm is only one possible measure of metacognition and that its contribution, though limited, may still be of value. Next we will consider some of the alternative methodologies currently in use and offer some suggestions for improving future comprehension monitoring research.

What can be done to further our understanding of comprehension monitoring? One answer (Brown, 1978; Kuhn, 1974) is to use a variety of tasks. An example is a paradigm that has been used by Brown and her associates (Brown & Smiley, 1977; Brown, Smiley, & Lawton, 1978). The subjects are asked to read or listen to a passage after which a recall measure is taken. They then are told to study in preparation for another recall test. Sometimes they are told to pick the retrieval cues they think will be most helpful. The data consists of two sets of recall and any notes or cues the subjects utilize. These data can be analyzed in such a manner as to provide the link between process and performance measures which are so important in interpreting strategy use and effectiveness (Ryan, in press).

Another related suggestion for improving comprehension monitoring research is for the experimenter to set the criterion for adequate comprehension. Both Markman (1977) and Baker (1979) noted that error detection improved when subjects were given explicit directions to find the problem. Therefore, it is important that we study comprehension monitoring under a variety of reading purposes before making any general statements about children's metacognitive abilities. The relationship between the reader's purpose and comprehension monitoring is too important to be left to chance.
There are several alternative comprehension monitoring paradigms that do not rely on the use of probes or other introspective verbal reports. This is important because performance measures, rather than introspective verbal reports, seem to be especially promising. Clay (1973), for example, has studied the self-correction of spontaneous errors by young children during oral reading. She found that the self-correction rate was more closely related to reading achievement scores in the first three years of instruction than either reading readiness scores or intelligence.

Danks et al. (Note 5) also used oral reading as the on-line measure. The hypothesis in that study was that different kinds of embedded errors would affect different components of the comprehension process. They found, for example, that embedded errors which violated syntactic and semantic constraints disrupted oral reading sooner than did semantic errors or logical inconsistencies. Results such as these may offer insights into the role of metacognition in an interactive model of reading. These results also indicate the importance of adequately specifying the kinds of errors used.

The two preceding paradigms have both used oral reading as the on-line measure. Although oral and silent reading may be similar in many respects, it is unsafe to assume they are identical (Danks & Fears, in press). One method, yet to be tried, is the use of eye movement technology in studying the effects of different kinds of embedded errors. Studies (McConkie, Note 6) are being planned which will extend the work of Danks and his colleagues. The goal is to develop a timeline representing the different latencies of the different kinds of embedded errors in an attempt to get separate measures of bottom-up and top-down processes in comprehension.
Another methodological approach is exemplified by the work of DiVesta, Hayward, and Orlando (1979). Their work is based on the assumption that effective strategy use reflects metacognitive skill. They devised a cloze test which measured differences in the reader's use of running and subsequent text. Their results indicated that there are developmental and ability differences in the strategies used in searching for needed information. Although these results are not surprising, the methodological approach is promising.

Each methodology has its own contribution to make and its own limitations to overcome. Thus, if a variety of experimental tasks and measures are used, more confidence can be placed in the data and any generalizations drawn from them. One such generalization that can be safely drawn now is that most readers are capable of exhibiting a wide variety of metacognitive abilities. However, many of these behaviors are extremely subtle, and the major difficulty lies in accurately measuring them. Since it is clear that most readers can monitor their comprehension, the major questions now involve specifying how and under what conditions they do so.
Reference Notes


6. McConkie, G. *Personal communication, October 17, 1979.*
References


Comprehension Monitoring

27


Glass, G. Students' misconceptions concerning their reading. *The Reading Teacher,* 1968, 21, 765-768.


Johns, J., & Ellis, D. Reading, kids tell it like it is. *Reading World,* 1976, 16, 115-128.


Kuhn, D. Inducing development experimentally: Comments on a research paradigm. *Developmental Psychology,* 1974, 10, 590-600.


Comprehension Monitoring


Appendix:

Pictures and Stories
Story 1

The view was breathtaking. From our seats we could see the crowd below. Everything looked extremely small from such a distance, but the colorful uniforms could still be seen. At first there was a great deal of activity. There seemed to be lots of children as well as adults and they all seemed to be looking for a place with a good view. The circus is always such an exciting place to visit. We all stood up and cheered when the music started. My mother's best dish fell off the dinner table and broke. We took many pictures of the setting and the crowd. Everyone was very friendly and we were glad that we had the chance to come.
Story 2

It was the best thing we ever saw. There were hundreds of people everywhere. Some were riding and some were walking. Things were going on all over the place. Some people were looking for places to buy food and drinks. There are so many things to do at the circus. After we bought our food we tried to find some good seats. It seemed like the big ship would sink in the rough sea. It was so noisy where we were that we could barely near the music. I could see some animals over on the side. We took lots of pictures of everything.
Story 3

When we arrived there were still plenty of seats. We had gotten there just in time because soon every seat was filled. I saw many people that I knew and I waved at some of them. All of a sudden things started to happen. The people in front of me stood up and I couldn't see a thing. All I could do was stand up and look at the church windows. Everybody finally sat down when the music was over. The teacher said it was time for us to put the toys away. My parents had been coming here for years but this was my first visit. I enjoyed it enough to come back again and again.
Story 4

It was very quiet when I sat down. The man in front began to talk. From where I sat I could barely hear him. I looked around at all the people. They were all dressed up in their best clothes but nobody was smiling. This was the first time I had ever been to my uncle's church. This was very different from anything I had ever seen before. When the batter hit a homerun, we all cheered. Everybody looked restless and some people began to yawn. When the man up front had finished speaking, the people got up to leave. Where I was sitting was very uncomfortable so I was glad it was over at last.
Table 1
Mean Performance Scores of High and Low Ability Students
With and Without Preparation

<table>
<thead>
<tr>
<th></th>
<th>Preparation</th>
<th>No Preparation</th>
<th>Row Means</th>
</tr>
</thead>
<tbody>
<tr>
<td>High Ability</td>
<td>3.03</td>
<td>3.17</td>
<td>3.1</td>
</tr>
<tr>
<td>Low Ability</td>
<td>5.79</td>
<td>5.62</td>
<td>5.71</td>
</tr>
<tr>
<td>Column Means</td>
<td>4.41</td>
<td>4.4</td>
<td>--</td>
</tr>
</tbody>
</table>
Table 2

Mean Performance Scores
across Stories and Treatment Conditions

<table>
<thead>
<tr>
<th>Preparation</th>
<th>Story 1</th>
<th>Story 2</th>
<th>Story 3</th>
<th>Story 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Preparation</td>
<td>1.55</td>
<td>4.55</td>
<td>9.67</td>
<td>3.11</td>
</tr>
<tr>
<td></td>
<td>(1.21)</td>
<td>(4.89)</td>
<td>(4.18)</td>
<td>(2.93)</td>
</tr>
<tr>
<td>No Preparation</td>
<td>5.44</td>
<td>4.44</td>
<td>4.54</td>
<td>4.1</td>
</tr>
<tr>
<td></td>
<td>(4.95)</td>
<td>(4.72)</td>
<td>(4.59)</td>
<td>(4.18)</td>
</tr>
</tbody>
</table>

Note: Standard deviations are shown in parentheses.
Table 3
Mean Performance Scores of High and Low Ability Students for Story 1

<table>
<thead>
<tr>
<th>Preparation</th>
<th>High Ability</th>
<th>Low Ability</th>
<th>Row Means</th>
</tr>
</thead>
<tbody>
<tr>
<td>Prep</td>
<td>1.2</td>
<td>1.8</td>
<td>2.1</td>
</tr>
<tr>
<td>No Prep</td>
<td>3</td>
<td>7.8</td>
<td>4.6</td>
</tr>
<tr>
<td>Column Mean</td>
<td>1.5</td>
<td>5.2</td>
<td>--</td>
</tr>
</tbody>
</table>
### Table 4

Mean Performance Scores of High and Low Ability Students for Story 2

<table>
<thead>
<tr>
<th>Preparation Type</th>
<th>High Ability</th>
<th>Low Ability</th>
<th>Row Means</th>
</tr>
</thead>
<tbody>
<tr>
<td>Preparation</td>
<td>1.8</td>
<td>6.8</td>
<td>2.9</td>
</tr>
<tr>
<td>No Preparation</td>
<td>4</td>
<td>4.8</td>
<td>5.8</td>
</tr>
<tr>
<td>Column Means</td>
<td>4.3</td>
<td>4.4</td>
<td>--</td>
</tr>
</tbody>
</table>
Table 5

Mean Performance Scores of High and Low Ability Students for Story 3

<table>
<thead>
<tr>
<th></th>
<th>Preparation</th>
<th>No Preparation</th>
<th>Row Means</th>
</tr>
</thead>
<tbody>
<tr>
<td>High Ability</td>
<td>7.3</td>
<td>4</td>
<td>5.7</td>
</tr>
<tr>
<td>Low Ability</td>
<td>11.6</td>
<td>4.8</td>
<td>8.2</td>
</tr>
<tr>
<td>Column Means</td>
<td>9.5</td>
<td>4.4</td>
<td>--</td>
</tr>
</tbody>
</table>
Table 6
Mean Performance Scores of High and Low Ability Students for Story 4

<table>
<thead>
<tr>
<th></th>
<th>Preparation</th>
<th>No Preparation</th>
<th>Row Means</th>
</tr>
</thead>
<tbody>
<tr>
<td>High Ability</td>
<td>2.3</td>
<td>2.6</td>
<td>2.5</td>
</tr>
<tr>
<td>Low Ability</td>
<td>3.8</td>
<td>5.3</td>
<td>4.6</td>
</tr>
<tr>
<td>Column Means</td>
<td>3.1</td>
<td>3.95</td>
<td>--</td>
</tr>
</tbody>
</table>
Table 7
Analysis of the Efficiency of the Interview Probes

<table>
<thead>
<tr>
<th>Score</th>
<th>Frequency</th>
<th>Probe #</th>
<th>Probe</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>33</td>
<td>2</td>
<td>Target hit during reading</td>
</tr>
<tr>
<td>2</td>
<td>15</td>
<td>3</td>
<td>Any comments?</td>
</tr>
<tr>
<td>3</td>
<td>2</td>
<td>4</td>
<td>Do you have any questions?</td>
</tr>
<tr>
<td>4</td>
<td>3</td>
<td>5</td>
<td>What did you think about the story?</td>
</tr>
<tr>
<td>5</td>
<td>5</td>
<td>6</td>
<td>Did everything make sense?</td>
</tr>
<tr>
<td>6</td>
<td>0</td>
<td>7</td>
<td>Could fifth graders understand everything?</td>
</tr>
<tr>
<td>7</td>
<td>0</td>
<td>8</td>
<td>Could you answer questions about it?</td>
</tr>
<tr>
<td>8</td>
<td>1</td>
<td>9</td>
<td>Would you change anything in the story?</td>
</tr>
<tr>
<td>9</td>
<td>0</td>
<td>10</td>
<td>Does everything sound all right?</td>
</tr>
<tr>
<td>10</td>
<td>5</td>
<td>11</td>
<td>Are there any sentences that don't fit?</td>
</tr>
<tr>
<td>11</td>
<td>16</td>
<td>12</td>
<td>Tell me what the story was about.</td>
</tr>
<tr>
<td>12</td>
<td></td>
<td></td>
<td>No hit at all</td>
</tr>
</tbody>
</table>
Figure Captions

Figure 1. Treatment effects within and across groups by testing sessions.

Figure 2. Reading ability by testing session.
No. 1: Durkin, D. *Comprehension Instruction—Where are You?*, October 1977. (ERIC Document Reproduction Service No. ED 146 566, 14p., PC-$1.82, MF-$0.83)


No. 4: Jenkins, J. R., & Pany, D. *Teaching Reading Comprehension in the Middle Grades*, January 1978. (ERIC Document Reproduction Service No. ED 151 756, 36p., PC-$3.32, MF-$0.83)

No. 5: Bruce, B. *What Makes a Good Story?*, June 1978. (ERIC Document Reproduction Service No. ED 158 222, 16p., PC-$1.82, MF-$0.83)

No. 6: Anderson, T. H. *Another Look at the Self-Questioning Study Technique*, September 1978. (ERIC Document Reproduction Service No. ED 163 441, 19p., PC-$1.82, MF-$0.83)


No. 8: Collins, A., & Haviland, S. E. *Children's Reading Problems*, June 1979. (ERIC Document Reproduction Service No. ED 172 188, 19p., PC-$1.82, MF-$0.83)

No. 9: Schallert, D. L., & Kleiman, G. M. *Some Reasons Why Teachers are Easier to Understand than Textbooks*, June 1979. (ERIC Document Reproduction Service No. ED 172 189, 17p., PC-$1.82, MF-$0.83)

No. 10: Baker, L. *Do I Understand or Do I not Understand: That is the Question*, July 1979. (ERIC Document Reproduction Service No. ED 174 948, 27p., PC-$3.32, MF-$0.83)


No. 3: Goetz, E. T. *Sentences in Lists and in Connected Discourse*, November 1975. (ERIC Document Reproduction Service No. ED 134 927, 75p., PC-$4.82, MF-$0.83)

No. 4: Alessi, S. M., Anderson, T. H., & Biddle, W. B. *Hardware and Software Considerations in Computer Based Course Management*, November 1975. (ERIC Document Reproduction Service No. ED 134 928, 21p., PC-$1.82, MF-$0.83)


No. 8: Mason, J. M. *Questioning the Notion of Independent Processing Stages in Reading*, February 1976. (Journal of Educational Psychology, 1977, 69, 268-297)


No. 15: Schwartz, R. M. *Strategic Processes in Beginning Reading*, November 1976. (ERIC Document Reproduction Service No. ED 134 937, 19p., PC-$1.82, MF-$0.83)

No. 16: Jenkins, J. R., & Pany, D. *Curriculum Biases in Reading Achievement Tests*, November 1976. (ERIC Document Reproduction Service No. ED 134 938, 24p., PC-$1.82, MF-$0.83)


No. 20: Kleiman, G. M. *The Effect of Previous Context on Reading Individual Words*, February 1977. (ERIC Document Reproduction Service No. ED 134 941, 76p., PC-$6.32, MF-$0.83)


No. 105: Ortony, A. *Beyond Literal Similarity*, October 1978. (ERIC Document Reproduction Service No. ED 166 635, 58p., PC-$4.82, MF-$0.83)


No. 115: Gearhart, M., & Hall, W. S. *Internal State Words: Cultural and Situational Variation in Vocabulary Usage*, February 1979. (ERIC Document Reproduction Service No. ED 165 131, 66p., PC-$4.82, MF-$0.83)


No. 120: Canney, G., & Winograd, P. *Schemata for Reading and Reading Comprehension Performance*, April 1979. (ERIC Document Reproduction Service No. ED 169 520, 99p., PC-$6.32, MF-$0.83)


No. 124: Spiro, R. J. *Etiology of Reading Comprehension Style*, May 1979. (ERIC Document Reproduction Service No. ED 170 734, 21p., PC-$1.82, MF-$0.83)


No. 147: Stein, N. L., & Goldman, S. Children's Knowledge about Social Situations: From Causes to Consequences, October 1979.


