PRODUCTION NOTE

University of Illinois at
Urbana-Champaign Library

No. 136 is missing
Technical Report No. 137

A SENTENCE VERIFICATION TECHNIQUE
FOR MEASURING READING COMPREHENSION

James M. Royer
University of Massachusetts, Amherst

C. Nicholas Hastings and Colin Hook
University of Illinois at Urbana-Champaign

August 1979

Center for the Study of Reading
Technical Report No. 137

A SENTENCE VERIFICATION TECHNIQUE FOR MEASURING READING COMPREHENSION

James M. Royer
University of Massachusetts, Amherst

C. Nicholas Hastings and Colin Hook
University of Illinois at Urbana-Champaign

August 1979

The research in this article was completed while the first author was a visiting professor at the Center for the Study of Reading, University of Illinois, Urbana-Champaign. The Center is supported by the National Institute of Education under Contract No. US-NIE-C-400-76-0116.
This page is intentionally blank.
Abstract

Two studies are reported which assess whether a sentence verification technique for measuring reading comprehension is sensitive to reading difficulty of text. Fifth and sixth grade students and fourth and sixth grade students read text passages two grade levels below reading level, at reading level, and two grades above reading level, and then responded to original, paraphrase, meaning change, and distractor sentences. Analysis of proportion correct and $d'$ scores indicated that the technique was sensitive to text difficulty. A number of possible uses and advantages of the sentence verification technique are discussed.
Measuring Reading Comprehension

A Sentence Verification Technique for Measuring Reading Comprehension

For the most part, techniques for measuring reading comprehension are independent of any theoretical considerations of reading as a psychological process. The two most popular techniques involve either having a reader examine a text passage and then respond to multiple choice questions, or having the reader fill in the blanks in a passage from which words have been deleted (the cloze technique). These techniques have evolved more as a function of pragmatism than of theory, and the continued use of the techniques is dependent more on psychometric criteria than on theoretical considerations.

This paper reports two studies using a technique for measuring reading comprehension which is based explicitly on theoretical notions about reading comprehension as a psychological process. Briefly, these notions are that language comprehension is a constructive process in which an incoming linguistic message, the recipient's prior knowledge, and the environmental context in which the message is received all interact to determine the nature of the comprehended message (e.g., Bransford & McCarrell, 1974; Kintsch, 1974; Kintsch & van Dijk, 1978; Royer & Cunningham, 1978; Smith, 1971). Moreover, the memorial representation of this comprehended message is in a form which preserves the meaning of the message but not its surface structure (Sachs, 1967; 1974).

Given the above perspective, one way to determine if a sentence has been comprehended is to establish whether this operation has been preferred,
i.e., whether the sentence is, in fact, represented in memory in a form which preserves the meaning, but not necessarily the surface structure of the message. We have developed a variation of a sentence verification technique which is designed to do precisely this. In our studies students read text segments which vary in difficulty, and then are presented with four types of test sentences which either are the same or mean the same as an original sentence, or have a meaning different from that of the original sentence. The student's task is to distinguish between these two classes of test sentences. If student performance on the task were to vary systematically as a function of text difficulty, we would interpret this as one source of evidence supporting the conclusion that the test is measuring reading comprehension.

**Experiment 1**

**Method**

*Design, subjects, and reading materials.* The intent of the study was to determine if our sentence verification technique could detect differences in comprehension when students were reading material of varying difficulty level. The study was designed such that each subject read and was examined on material which was at a difficulty level well below the reading level of the average reader, at the average reading level, and well above the average reading level.

The subjects were 21 fifth graders and 33 sixth graders from a small school district in Central Illinois. At our initial meeting with the teachers of the two classes we explained our technique and showed
them sample test materials. We then asked the teachers to select text samples which were about two grade levels below (in difficulty) the level at which the average child in the class was reading, right at the level where the average child was reading, and two grade levels above where the average child was reading. Furthermore, these samples were to be ones not seen by the children before. Thus, we had text samples which were approximately at the third, fourth, fifth, sixth, seventh, and eighth grade levels of difficulty.

Test materials. Using the text samples identified by the teachers, we first divided the texts into two segments, each of which was twelve sentences long. We then developed four test sentences for each of the sentences which appeared in the original text. These sentences were: (1) the original sentence exactly as it had appeared in the text, (2) a paraphrase of the original sentence in which as many words as possible were changed without altering the meaning or the syntactical structure of the original sentence, (3) a meaning change sentence in which one word in the original sentence was altered such that the meaning of the sentence was changed, and (4) a distractor sentence which was consistent with the general theme of the passage, but was unrelated to any of the original sentences. In addition, the distractor sentence was written (intuitively) to be of the same length, syntactical structure, and difficulty level as the original sentence. Sample items for third and eighth grade materials are presented in Table 1.
After developing the four types of sentences for each sentence in the original text, we stored the sentences, and the words "old" and "new" underneath each sentence, on computer tape. A program was then written which randomly generated (without replacement) sixteen sentence tests based on each of the twelve sentence segments of original text. The single restriction to the random rule was that the first eight test sentences in a test came from the first six sentences of the original text (this was to control for recall from short term memory). Using this program, we generated a different test form for each of the students who participated in the study.

The actual test booklets consisted of the following items: (1) a cover sheet which was color coded for easy identification; (2) a short, easily understandable context setting paragraph, the purpose of which was to provide a schematic reference for reading the text samples; (3) a text sample twelve sentences in length; (4) the sixteen item test based on the text sample immediately preceding it; (5) a second context setting paragraph; (6) a second text sample at the same level of difficulty as the first; and (7) a second sixteen item test based on the second text sample.

**Procedure.** The first day of the study was devoted to training the students on how to perform on the tests. At the beginning of the training period booklets were handed out which were similar to those used in the actual study. We then explained that the purpose of our study was to try
out a new technique for measuring reading comprehension and that the technique involved reading a segment of text and then judging whether test sentences were "old" or "new." It was then indicated that old sentences were sentences that were the same or meant the same as an original sentence, whereas new sentences were either different in meaning or unrelated to an original sentence. The students were then instructed to turn to the first page of their booklets which contained a context setting paragraph. After reading this paragraph, the students turned the page to a twelve sentence text sample (estimated to be well below the average reading level of both classes) we had made up for the purposes of training. The instructions on this page were to read the text slowly and carefully twice. After the children had finished reading, we asked them to turn the page. The page they turned to contained four sentences—one example of each of the four test sentence types—based on the first sentence to appear in the original text. At the same time we flashed on a screen (via overhead projector) the same four sentences. We then went through each of the four sentences, indicating which of the sentences would be marked old and new (and why), and urging the children to turn to the original text to compare the original sentence to the test sentences. After going through the four sentences, we asked for and responded to questions, and then instructed the children to turn to the next page. This page contained the test sentences for the second sentence appearing in the original text, and as the children turned to the page, we projected the sentences onto the
screen in the front of the class. We then asked the children to shout out how the first sentence would be responded to. The children were then given the correct answer and told why it was correct, and we then went on to the next sentence. This procedure continued until we had asked for responses and questions and given feedback on all the sentences.

The process described above continued until we had examined the test sentences based on the first six sentences in the original text. During this procedure it was noted that very few children failed to correctly respond with the correct classification for each of the test sentences. After completing the above described procedure for the six original text sentences, we asked the children to turn to a page which contained only one of the test sentence types (randomly drawn but the same for each child) for each of the remaining six original text sentences. Under each of the sentences were the words "old" and "new," and the children were instructed to circle the correct classification for each of the six test sentences and to not look back to the original text. After all of the children had completed their ratings, we projected the sentences on the screen and gave the correct classification for each sentence along with the reason for the classification. While this explanation was being given, two experimenters walked up and down the aisles giving personal attention to any student who had an incorrect answer.

After a call for, and a response to, general questions, a second test booklet was handed out. This booklet contained the seven types of material described in the test materials section. The text was based on material judged to be at an average level of difficulty. After general
instructions to read the context paragraphs, to read the text slowly and carefully twice, and to complete the test sentences in the manner practiced, the children read through the next two samples, and completed the two sets of test sentences at their own pace. During this period we were available to answer any questions which might arise. The purpose of this second test booklet was to give the children practice under actual test conditions.

On the next day we conducted the actual study. Each child was given an envelope containing a set of general instructions and three test booklets, each of which had a different color cover sheet. These test booklets had been made up such that each of the possible combinations of ordering the three difficulty levels of text was nearly equally represented by each color. After reading the instructions aloud, and having the children read along with us, we instructed the children to pull out of their envelopes a particular colored test booklet. They were then instructed to follow the instructions in the booklets, and to raise their hands when they had finished. When a child's hand was raised, an experimenter checked the booklet to see that all of the test sentences had been responded to, replaced that booklet in the envelope, and gave the child the appropriate next booklet. This procedure continued until all of the children had completed all three of the test booklets.

**Dependent variables.** The task in this study can be considered as a two choice discrimination problem. Typically, the data of interest in such tasks is the proportion of correct classifications, and we will analyze proportions in this study. There is a problem, however, with using
proportion correct as an index of classification accuracy: Response accuracy is confounded with criterion differences. Imagine, for example, that one of our students was willing to classify a sentence as "old" when he/she had a slight feeling that the sentence was the same as one appearing previously. Imagine that another student classified a sentence as old only when he/she was absolutely certain that the sentence was the same as an original text sentence. The two students might be identical with respect to their ability to correctly classify the sentences, but their performance would differ markedly because of their criterion differences.

Fortunately, there is a way to separate response accuracy from the criterion for a response. The technique involves the use of signal detection analysis (e.g., Swets, Tanner, & Birdsall, 1961), in which a cutoff parameter ($\beta$) and a signal sensitivity ($d'$) parameter are separately computed. The $d'$ parameter is a criterion free measure of response accuracy (see Healy & Kubovy, 1978 for evidence of this) and will be of central interest in this study. A $d'$ score can be viewed as a $Z$ score indicating the likelihood that the individual is responding at chance. Thus, a $d'$ of 3.0 is a very high score indicating performance substantially above chance. In contrast, a score near zero would indicate the individual is guessing when classifying the sentence as old or new.

Results

Analyses of variance. Both the proportion of correct scores and the $d'$ scores were analyzed using a 2(grade level, 5 and 6) x 3(low, average, or
high difficulty level text material) analysis of variance with grade level a between subjects factor and difficulty level a within subjects factor. The means for both dependent variables are presented in Table 2. These means are based on 96 responses to test sentences (approximately 32 to easy material, 32 to on grade material, etc.) for each individual.

The analysis of variance for the $d'$ data indicated that difficulty level was a significant source of variance, $F(1,52) = 5.3$, $p < .05$, and the interaction between grade level and difficulty level, $F(2,104) = 4.7$, $p < .05$, were also significant.

Of additional interest was performance on the four different types of test sentences. Proportion correct on the four types broken down by grade and difficulty level is presented in Table 3. An analysis of variance on these data indicated that in addition to the previously mentioned significant effects, sentence type was also a reliable source of variance, $F(3,156) = 70.8$, $p < .01$.

Correlations. The students who participated in our study had recently taken the achievement series from the Iowa Tests of Educational Development (ITED), and we performed a number of correlational analyses using these scores and scores from our sentence verification test. These
correlations should be taken as merely suggestive since our sample size was too small to provide stable correlations and since we used percentile scores from the ITED rather than raw scores. Given these cautions, the correlations between the ITED reading comprehension subscore and the sentence verification overall proportion correct and overall d' scores were, respectively, .73 and .58. Another set of correlations of interest were those between the Short Test of Educational Ability scores (a general ability subscale of the ITED), ITED reading comprehension scores, and the proportion correct and d' scores from the sentence verification test. As is generally the case with standardization tests, the relationship between the ability score and the ITED reading comprehension score was relatively high and positive (r = .79). In comparison, the relationship between the ability scores and the proportion correct and d' scores from the sentence verification tests were somewhat lower (r = .53 and .41, respectively).

**Experiment 2**

When we initially planned our study, we had intended to have a sample of fourth graders and a sample of sixth graders so that we could have children in the two grade levels reading and being tested on some of the same materials. That is, our on grade material for the fourth graders would be below grade material for the sixth graders, and above grade material for the fourth graders would be on grade material for the sixth graders. Because of the unavailability of a group of fourth graders, we...
were not able to carry out this plan in our initial study. Soon after completing the first study, however, we were able to find a sample of fourth and sixth graders, and we then conducted our study as originally intended.

Method

Subjects and materials. The subjects were 20 fourth graders and 22 sixth graders from a medium sized school district in Central Illinois. After developing materials at about the second grade level for use as below grade materials for the fourth graders, we asked the teachers of the two classrooms to examine our materials to determine if they would be appropriate for their students. The teachers agreed that the materials which were approximately at the second, fourth, sixth, and eighth grade levels of difficulty would be appropriate for use with their students.

Design, procedure, dependent variables. With two exceptions, all aspects of the study were identical to those reported for Experiment 1. The first aspect that was different was that we reprogrammed our item generation routine to construct sixteen-sentence tests to consist of exactly four original items, four paraphrase items, four meaning change items, and four distractor items. The second change was that rather than generating a different test form for each student, we generated three test forms which were used with equal frequency in the study.
Results

Analyses of variance. The means for both the proportion correct and the $d'$ dependent variables are listed in Table 4. These data were analyzed in a $2(\text{grade level, 4 or 6}) \times 3(\text{low, average, or high difficulty level})$ analysis of variance with the difficulty factor being a within subjects variable.

The analysis of the proportion correct data indicated that both the grade level effect ($F = .25$) and the interaction effect ($F = 1.4$) were nonsignificant, but that the difficulty effect was reliable, $F(2,80) = 18.8, p < .01$. The analysis of the $d'$ scores revealed a similar pattern in that the grade level effect ($F = 1.5$) and interaction ($F = 2.0$) were nonsignificant whereas the effect for difficulty level was significant, $F(2,80) = 68.2, p < .01$.

As noted earlier, the main reason for wanting groups of subjects two grades apart was to allow a comparison between the groups on the same material. It is instructive to examine the means relevant to this comparison for both the proportion correct and the $d'$ dependent variables. Note first that the proportion correct data are disappointing in that the fourth grader's on grade performance is equal to the performance for the sixth graders on the same material. Since this is the same material, and since sixth graders presumably read with greater comprehension than do fourth graders, it should not be the case that they perform at the
same level. However, now examine the $d'$ data. These scores are much more consistent with expectations in that the sixth graders are performing much better than the fourth graders when both groups are reading the same material. For those readers not familiar with $d'$ scores, it should be pointed out that the difference between a mean score of 2.49 for the sixth graders and 1.38 for the fourth graders is as great as the difference between $Z$ scores of the same magnitude. The implications of these comparisons will be discussed later in the paper.

The final analysis of variance was an analysis using proportion correct scores which included type of test sentence as a variable. This analysis indicated that type of item was a significant source of variance, $F(3,120) = 40.1, p < .01$, with the pattern of means being very similar to those reported in Experiment 1.

Correlations. The subjects in the study had recently taken the Stanford Achievement Test (SAT) and several of the scores from this test were correlated with the proportion correct and $d'$ scores from our sentence verification test. Again, these correlations should be interpreted with caution because of the small sample size and the use of percentile scores in computing the correlations. The correlations between the SAT reading comprehension score and the proportion correct and $d'$ score from the sentence verification test were .49 and .50, respectively.

Discussion

The primary purpose of the studies reported in this paper was to determine if a sentence verification technique would detect differences
in performance when subjects read materials of varying difficulty. The analyses of variance reported for both studies indicate that the technique was sensitive to differences in text difficulty. The analyses of both the \( d' \) measure and the proportion correct measure found that difficulty level of the materials was a highly significant source of variance.

An important methodological point can be found in the comparison between the proportion correct and the \( d' \) dependent variables. In every case the \( d' \) measure provided a more powerful (in the sense of detecting deviation from the null hypothesis) analysis than did the proportion correct measure. Furthermore, in the comparison of performance of fourth and sixth graders on the same material in Experiment 2, the proportion correct measure failed to detect a difference whereas the \( d' \) measure revealed a very large difference between the two grade levels. These comparisons suggest that the criterion for making a response varies as a function of the skill level of the reader and the difficulty level of the text. This in turn stresses the importance of a criterion-free measure of response accuracy.

In addition to providing a criterion-free measure of response accuracy, \( d' \) scores derived from a sentence verification test also possess another quality which could make them extremely useful as an index of text comprehensibility. A \( d' \) score is an absolute index in the sense that it reflects deviation from chance performance independent of the nature of the text material. This quality could prove useful in a setting (e.g., industrial or military) where instructional materials were being developed.
with a particular student population in mind. The text developers could set a criterion for text comprehensibility (say a $d'$ of 2.0), and materials could be run through a test-revision cycle until that criterion was reached.

It is also possible that the scores from the sentence verification technique have diagnostic utility. The four types of test sentences have been included because they are sensitive to different hypothesized styles of reading. Competent readers should have little difficulty with any of the sentence types. Presumably, they have transformed each original sentence into a form which preserves the meaning but not the surface structure of the sentence. The reader should then be able to identify original and paraphrases as sentences which mean the same thing as an original text sentence, and to identify meaning change sentences and distractors as being different in meaning from the original text sentence.

In contrast to the able reader, a less capable reader should have difficulty with the task. Imagine, for example, a word by word reader who has difficulty grasping the meaning of the words being read. The reader might perform perfectly on the original and distractor sentences (because the words are either exactly the same or entirely different), but perform very poorly on the paraphrase (new words, old meaning) and meaning change sentences (old words, new meanings).

Another possible type of reader would be one who relied heavily on his or her world knowledge while reading, to the point where the words were not carefully attended to. This reader might perform well
on original and paraphrase sentences, but miss the change in meaning change sentences. Moreover, it is possible that this reader would respond "old" to distractor sentences since they are consistent with the theme of the original passage.

Another obvious use for the sentence verification technique is as a measure of comprehension in experimental research. Whereas techniques like free recall of text have apparent positive qualities, they also have several negative qualities. For example, they are difficult to use with younger children, and the researcher is frequently concerned that much of the variability in recall among subjects is associated with motivational and emotional (e.g., shyness) factors. It is likely that several of these sources of variability may be lessened with the sentence verification technique. Furthermore, the technique is usable with a wide variety of texts.

Another aspect of the sentence verification technique which should be noted is that it does not require inferential reasoning. A subject can respond to a test sentence by comparing that sentence to the memorial representations of the original sentences. In contrast, the questioning techniques and the cloze techniques used on standardized tests are heavily influenced by inferential reasoning and general world knowledge (Pyrczak, 1972; Royer & Cunningham, 1978; Tuinman, 1973-74). This influence is, in turn, likely to result in a substantial relationship between reading comprehension test scores and scores in general ability tests (cf. the correlations in Experiment 1). Future research will have to
establish the extent to which performance on the sentence verification test is influenced by general ability factors (e.g., reasoning ability, world knowledge). But at this point it seems likely that the scores will be less influenced by ability factors than are other popular techniques (cf. the correlations in Experiment 1).

Even though the results of the two studies reported in this paper are encouraging, it is obvious that much research remains to be done. Rules for generating the various types of test sentences need to be developed rather than doing this on an intuitive basis as we did. Also, it is necessary to develop text materials which have known comprehension properties rather than simply rely on the judgment of teachers. One way we plan to approach this problem is to develop materials using the text analysis procedure developed by Walter Kintsch at the University of Colorado (see Kintsch & Vipond, 1977; Turner & Greene, Note 1). Finally, large scale studies need to be conducted which would determine the extent to which the sentence verification technique is a reliable and valid measure of reading comprehension.
Reference Note

References


Table 1
Sample Sentences for Third and Eighth Grade Materials

<table>
<thead>
<tr>
<th>Sentence Type</th>
<th>Third Grade</th>
<th>Eighth Grade</th>
</tr>
</thead>
<tbody>
<tr>
<td>Original</td>
<td>One drift was piled against the garage door.</td>
<td>At length I came to know him more thoroughly than anyone else.</td>
</tr>
<tr>
<td>Paraphrase</td>
<td>One mound blocked the entrance to the garage.</td>
<td>After a time I knew him better than any other person.</td>
</tr>
<tr>
<td>Meaning Change</td>
<td>One drift was piled against the cellar door.</td>
<td>At length I came to know him less thoroughly than anyone else.</td>
</tr>
<tr>
<td>Distractor</td>
<td>If it kept snowing the schools would be closed next week.</td>
<td>The mesas of New Mexico are strewn with rocks and canyons.</td>
</tr>
</tbody>
</table>
Table 2
Mean Proportion Correct and \( d' \) as a Function of Grade Level and Text Difficulty

<table>
<thead>
<tr>
<th>Text Difficulty</th>
<th>Grade 5</th>
<th></th>
<th>Grade 6</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>( P(C) )</td>
<td>( d' )</td>
<td>( P(C) )</td>
<td>( d' )</td>
</tr>
<tr>
<td>Below level</td>
<td>.83</td>
<td>2.5</td>
<td>.79</td>
<td>1.84</td>
</tr>
<tr>
<td>On level</td>
<td>.77</td>
<td>1.33</td>
<td>.71</td>
<td>1.23</td>
</tr>
<tr>
<td>Above level</td>
<td>.70</td>
<td>1.24</td>
<td>.68</td>
<td>1.19</td>
</tr>
</tbody>
</table>
Table 3
Mean Proportion Correct as a Function of Grade, Text Difficulty, and Sentence Type

<table>
<thead>
<tr>
<th>Sentence Type</th>
<th>Grade 5 Below</th>
<th>Grade 5 On</th>
<th>Grade 5 Above</th>
<th>Grade 6 Below</th>
<th>Grade 6 On</th>
<th>Grade 6 Above</th>
</tr>
</thead>
<tbody>
<tr>
<td>Original</td>
<td>.93</td>
<td>.89</td>
<td>.80</td>
<td>.88</td>
<td>.86</td>
<td>.78</td>
</tr>
<tr>
<td>Paraphrase</td>
<td>.67</td>
<td>.54</td>
<td>.56</td>
<td>.72</td>
<td>.62</td>
<td>.65</td>
</tr>
<tr>
<td>Meaning Change</td>
<td>.74</td>
<td>.73</td>
<td>.54</td>
<td>.67</td>
<td>.52</td>
<td>.46</td>
</tr>
<tr>
<td>Distractor</td>
<td>.97</td>
<td>.90</td>
<td>.92</td>
<td>.88</td>
<td>.86</td>
<td>.84</td>
</tr>
</tbody>
</table>
Table 4

Mean Proportion Correct and $d'$ as a Function of Grade Level and Text Difficulty

<table>
<thead>
<tr>
<th>Text Difficulty</th>
<th>Grade 4</th>
<th></th>
<th>Grade 6</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>P(C)</td>
<td>d'</td>
<td>P(C)</td>
<td>d'</td>
</tr>
<tr>
<td>Below level</td>
<td>.88</td>
<td>2.98</td>
<td>.84</td>
<td>2.50</td>
</tr>
<tr>
<td>On level</td>
<td>.84</td>
<td>1.38</td>
<td>.81</td>
<td>1.36</td>
</tr>
<tr>
<td>Above level</td>
<td>.75</td>
<td>1.26</td>
<td>.77</td>
<td>1.27</td>
</tr>
</tbody>
</table>
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. 9: Schallert, D. L., & Kleiman, G. M. *Some Reasons Why Teachers are Easier to Understand than Textbooks*, June 1979.

No. 10: Baker, L. *Do I Understand or Do I not Understand: That is the Question*, July 1979.

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-$83)
No. 8: Mason, J. M. Questioning the Notion of Independent Processing Stages in Reading, February 1976. (Journal of Educational Psychology, 1977, 69, 288-297)
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-$83)


No. 26: Armbruster, B. B., Stevens, R. J., & Rosenshine, B. **Analyzing Content Coverage and Emphasis: A Study of Three Curricula and Two Tests**, March 1977. (ERIC Document Reproduction Service No. ED 136 238, 22p., PC-$1.82, MF-$1.82)


No. 37: Adams, M. J. **Failures to Comprehend and Levels of Processing in Reading**, April 1977. (ERIC Document Reproduction Service No. ED 145 410, 51p., PC-$4.82, MF-$1.82)


No. 130: Bruce, B.  *Analysis of Interacting Plans as a Guide to the Understanding of Story Structure*, June 1979.


