ILLINOIS

UNIVERSITY OF ILLINOIS AT URBANA-CHAMPAIGN

PRODUCTION NOTE

University of Illinois at
Urbana-Champaign Library
Large-scale Digitization Project, 2007.

Technical Report No. 271

THE READING GROUP:
AN EXPERIMENTAL INVESTIGATION OF A LABYRINTH

Richard C. Anderson, Jana Mason
and Larry Shirey

University of Illinois at Urbana-Champaign

February 1983

Center for the Study of Reading

TECHNICAL REPORTS

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

BOLT BERANEK AND NEWMAN INC.
50 Moulton Street
Cambridge, Massachusetts 02238

The National Institute of Education
U.S. Department of Education
Washington, D.C. 20208

CENTER FOR THE STUDY OF READING

Technical Report No. 271

THE READING GROUP: AN EXPERIMENTAL INVESTIGATION OF A LABYRINTH

Richard C. Anderson, Jana Mason
and Larry Shirey

University of Illinois at Urbana-Champaign

February 1983

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

Bolt Beranek and Newman Inc.
50 Moulton Street
Cambridge, Massachusetts 02238

The research reported herein was supported by the National Institute of Education under Contract No. US-NIE-C-400-76-0116. The authors gratefully acknowledge the cooperation of the third grade teachers and children from Urbana, Ill., Homewood, Ill., and Rantoul, Ill. who participated in this research. We also are pleased to acknowledge the important contributions made by Taffy Raphael, Paul Wilson, William May, William Tirre, William Nagy, and Alexis Freeman. Some of the material contained in this paper was presented in an address at the International Reading Association convention in Chicago, May 1982.

Abstract

Reported are two experiments with third graders in which a number of dimensions of reading instruction were investigated. The major findings: an emphasis on meaning produces better results than an emphasis on word identification; in groups receiving a word identification emphasis, but not a meaning emphasis, results depend upon instructional time; the child who is taking an active turn gets more from a lesson than the children who are following along; and the interestingness of the material is a major factor in performance, one that is much more important than readability.

The Reading Group: An Experimental Investigation of a Labyrinth

> There is properly no history; only biography. Every mind must know the whole lesson for itself, must go over the whole ground. What it does not see, what it does not live, it will not know. (Ralph Waldo Emerson, Essays, 1842/1945, p. 6)

The Great Debate (Chall, 1967) in beginning reading instruction is over the relative emphasis that ought to be given to decoding and meaning. The available evidence favors a decoding emphasis. It appears that reading programs that begin with explicit, direct instruction in spelling-sound correspondences are more successful than programs that rely on incidental learning of these correspondences (Pflaum, Walberg, Kanegianes, & Rasher, 1980). However, it is possible that programs that include a substantial amount of direct instruction in spelling-to-sound correspondences are successful for other reasons. Such programs typically are more structured, provide more systematic feedback, allocate more time to reading, and maintain higher levels of student engagement than meaning emphasis programs, whose advocates often believe that learning to read is a "natural process" in which it is unwise to intervene heavily (Goodman & Goodman, 1979). Thus, it can be argued that program evaluations and related teacher effectiveness research have underrepresented classrooms in which the instruction is both meaning-oriented and structured and systematic. There is at least one beginning reading program that features both a meaning orientation and systematic direct instruction, the Kamahamaha Early Education Project. It is thoroughly documented that this program achieves good results with at-risk minority children (Tharp, 1982).

If one were to grant that direct instruction designed to produce competence in fast, accurate word identification typically is best at the beginning, an important policy question would still remain. No one doubts that the eventual goal is for children to read with comprehension. The question is, therefore, at what point in a child's development of reading proficiency should the schools stop stressing word identification and begin placing predominant emphasis on meaning?

On the one hand, Venezky and Massaro (1979) have doubted that more instruction in word identification could ever be too much of a good thing. They concluded an article on the importance of rapid word recognition by saying they were no longer willing to agree, as they once did, with Chall (1967, p. 307) who they then quoted as follows:

> Once the pupil has learned to recognize in print the words he knows (because they are part of his speaking and listening vocabulary), any additional work on decoding on his part is a shear waste of time.

Similarly, Perfetti and Lesgold (1977, 1979) have summarized research that shows large differences between good and poor readers in speed of pronunciation of unfamiliar words and pronounceable nonwords. This evidence establishes beyond any reasonable doubt that good and poor readers differ in their understanding of spelling-to-sound correspondences, whereas, in their research, Perfetti and Lesgold were unable to find evidence that good and poor readers differ substantially in use of higher level text information. From this they concluded that reading programs should include even more practice than typically given in accurate word pronunciation, more drill in speeded word recognition, and more practice in immediate memory for the facts in stories and texts. However, as Perfetti and Lesgold

acknowledged, attempts to improve reading comprehension using speeded word drills have not yet proved very successful (Jenkins, Pany, & Schreck, 1978).

On the other hand, there is a cornucopia of evidence that could be cited to support a stress on meaning in the reading program. Indeed, the benefits of an emphasis on meaning observed in basic research are so strong and so consistent that it is possible, in the fashion of scientists of another generation, to proclaim an empirical law. It might be called the Law of Meaningful Processing and it can be formulated as follows: Other things being equal, people learn and remember more when conditions require them to understand the material. There is a problem in specifying exactly what is meant by terms such as "meaningful" and "understand" (Baddeley, 1978), but for the moment informal, ordinary language senses will do.

A great deal of the research illustrating the Law of Meaningful Processing has involved recognition or recall of lists of words. A study by Craik and Tulving (1975, Ex 1) is a good example, though convincing evidence from many earlier studies is available (e.g., Baddeley, 1966; Hyde & Jenkins, 1969). Craik and Tulving had college students make judgments about a series of words presented one at a time. The judgment tasks ranged from those that did not demand any meaningful processing to ones that demanded a lot. Four of the tasks will be described. The first required only processing of graphemic information. The question was, "Is the word in capital (or lower case) letters?"--TABLE. The second task required processing of phonemic information. The question was, "Does the word rhyme with (for instance) weight?"--CRATE. The third task involved judging whether the word named a member of a certain category. The question was, "Is the word a type of (for instance) fish?"--SHARK. The fourth task

required determining whether a word could be inserted in a certain sentence. The question was (for instance), "Would the word fit in the sentence, 'He met a _____ in the street'?"--FRIEND. The subjects thought the point of the experiment was to determine how quickly the judgments could be completed, but when they were finished a surprise test was given that required recognizing the words that had been judged and discriminating them from other words. Pooling over the words for which the right answers in the earlier judgment task had been "yes" and "no," the percentage of words recognized from the graphemic, phonemic, category, and sentence tasks were 16%, 57%, 78%, and 90%, respectively.

A number of experiments have studied the effects of meaningful processing on sentence learning and remembering. For illustration we will present several examples from our own research. Anderson and Hidde (1971) had college students rate either the pronouncibility or the image-evoking potential of a series of sentences and then gave a surprise recall test. Notice that creating an image certainly requires contact with meaning whereas a person can evaluate how easily a sentence rolls off the tongue without constructing a meaningful representation. Subjects who rated imagery recalled 65% of the sentences while those who rated pronouncibility recalled only 25%. Anderson, Goldberg, and Hidde (1971) and Kane and Anderson (1978) asked adult subjects to read aloud sentences such as "Elevators stop at every floor." Half of the time there was a blank in the place of the last word and the word that fit the blank was highly predictable. The hypothesis was that filling blanks would increase learning since a reader cannot supply a missing word in a sentence of the

type illustrated without actively bringing to mind a meaningful representation of the rest of the sentence, whereas pronouncing the same sentence as a whole can be done without accessing meaning. The results from several experiments have confirmed this hypothesis; subjects who filled blanks recalled about 10% more sentences on the average. Closer to instruction is a study by Anderson and Kulhavy (1972). College students read definitions of difficult, unfamiliar concepts, such as atavistic, cuprous, palliate, and xanthous, and then answered multiple choice items that required identifying examples of the concepts. As they read, half of the subjects were instructed to create and say aloud a sensible sentence containing each defined word. For instance, given the definition "Atavistic means reversion to a primitive type," one subject said, "Mayor Daley's politics are atavistic." The remaining subjects read each definition aloud three times, a task which took about the same amount of time as composing a sentence. After one exposure to the definitions, subjects who had composed sentences got 65% of the test items correct whereas subjects who had orally repeated the definitions got 44% correct.

Finally, research consistently indicates that conditions that promote meaningful processing facilitate learning from connected text. For instance, Schallert (1976) found that subjects given directions to rate paragraphs for clarity/ambiguity or for difficulty, two tasks that require evaluating meaning, performed substantially better on recall and recognition tests than subjects who counted four-letter words or counted number of personal pronouns, two tasks that do not require much contact with meaning. Some of the studies with text have obvious practical as well as theoretical implications. Watts and Anderson (1971) asked high school students to

answer a question after reading each of several passages explaining psychological principles. The question involved identifying an example of the principle or the name of the psychologist associated with the principle. Students whose questions required them to apply the principles to new examples performed significantly better on a later test than all other students, including students who received otherwise identical questions that repeated examples described in the text. Students who answered name questions performed worst of all, poorer even than students who read without questions. Watts and Anderson argued that application-to-new-example questions induced students to process the passages in a deep, meaningful manner. In the same vein, Glover and his associates (e.g., Glover, Bruning, & Blake, 1982) have done a number of experiments showing better learning and remembering of the information in a text when students are required to formulate logical extensions of the text, paraphrase the text, or judge inferences from the text, all activities that require meaningful processing.

While the facts are clear, just why meaningful processing facilitates learning and remembering has been a matter of controversy. In an influential paper, Craik and Lockhart (1972) proposed what they called a "levels-of-processing" theory. The essential idea was that every level of analysis of written language--graphemic, phonemic, semantic--leaves a trace in memory and the "deeper" the analysis the more enduring the trace. This theory has been criticized for a variety of reasons, notably because it is vague, indeed, little more than a restatement of the facts (Baddeley, 1978). In subsequent papers, Craik and his colleagues have revised their formulation and now speak of "distinctiveness" (Jacoby, Craik, & Begg, 1979) and "rich semantic elaboration" (Craik & Tulving, 1975).

The term "semantic elaboration" refers to the subject's tracing connections between the presented material and background knowledge. While research has confirmed that elaboration plays a role (Ross, 1981), it is now also clear that the sheer amount of elaboration is not the key. Stein and Bransford (1979) found that subjects were slightly worse at recalling core sentences such as "The fat man read the sign" when the sentences were elaborated as in, "The fat man read the sign that was two feet high." In contrast, recall of the core sentences was substantially better when "precise elaborations," as in "The fat man read the sign warning of the thin ice," were added. A precise elaboration clarified the significance of concepts in the core sentence and indicated how the concepts fit together. An "imprecise elaboration," on the other hand, extended the core sentence in an arbitrary manner; it did nothing to clarify the core sentence or show how the constituent concepts related to one another.

A further insight into meaningful processing has been provided in research beginning with the studies of Anderson, Goldberg, and Hidde (1971) already reviewed: It is important for the reader to be involved in an active effort after meaning. Evidence confirming this point has been provided in a clever series of experiments by Auble, Franks, and Soraci (1979) who presented sentences such as "The party stalled because the wire straightened." Five seconds later a clue was presented that helped the subject figure out the sentence's meaning. The clue was "corkscrew" in this case. Subjects recalled more sentences when they received this arrangement than when they spent the same amount of time reading a sentence that already embodied the clue word, in this case, "The party stalled because the corkscrew wire straightened." The investigators concluded

that the results support the importance of an "aha" experience which entails the readers' making an active effort to move from a state of incomprehension to a state of comprehension.

To summarize, while there is still much to be learned, we know two of the boundary conditions that explain the inner workings of the Law of Meaningful Processing. First, learning is facilitated when information is elaborated in such a way that all of the information is integrated into a coherent representation. Second, learning is facilitated when the reader is actively involved in attempting to generate a coherent representation. Stating these fundamental points in the elegant language of the nineteenth century American philosopher, Ralph Waldo Emerson (1842/1945) "Every mind must know the whole lesson for itself, must go over the whole ground. What it does not see, what it does not live, it will not know" (p. 6).

It is a most important fact that young children and low ability persons are less likely than older or more able persons to engage spontaneously in meaningful processing. One of several studies pointing to this conclusion was done by Paris and Lindauer (1976). First and fifth graders listened to sentences such as "The workman dug a hole in the ground," or, alternatively, "The workman dug a hole in the ground with a shovel." Given the instrument word, such as "shovel," as the cue, the first graders recalled many more sentences when the instrument had been explicitly mentioned than when it had been left implicit. In contrast, fifth graders recalled almost as many sentences in which the instrument used to perform some action was unstated as they did sentences in which it was stated. These results appear to mean that the fifth graders spontaneously drew inferences that meaningfully integrated the information in the sentences with background knowledge, while

the first graders did not. In another experiment, first graders acted out sentences they heard. Under these conditions their performance on sentences with unstated instruments improved sharply, presumably because they were now meaningfully integrating the sentences with background knowledge. The evidence suggesting that young and low achieving children do not readily engage in meaningful processing provides support for an emphasis on meaning among young readers, and some grounds for the fear that a steady diet of word identification drill will produce children who are nothing but "word callers."

It is often said that classroom instruction is a "complex" process. Provided this is more than a code word for the belief that instruction is a mystery beyond the pale of empirical investigation, we take the view that instruction is complex to mean that interactions are expected among aspects of teaching method, setting, materials, and student characteristics. Very little instructional research has been designed so that it could illuminate interactions. Experimental investigations typically study the effects of at most one or two factors with everything else allegedly held constant. Systematic classroom observation studies generally attempt to measure many factors, but the data are often aggregated by classroom and are usually aggregated over weeks or even months of instruction during which each teacher has used different techniques and a variety of kinds of materials and each student has played various roles and engaged in various activities. If instruction truly is a complex process, then the trends that could emerge in aggregated data would only be a dim reflection of the interplay of forces at work.

In elementary school classrooms, the largest portion of children's reading instruction takes place when the teacher works on reading with small groups of students. Particularly in the first three or four grades, the children usually are divided into several groups according to ability. While one group works with the teacher, the others complete workbooks and exercise sheets at their seats. Though the sequence is not usually this complete or orderly (Mason, 1982), the children in the reading group are introduced to the new words in the day's basal reader story and may be helped to develop some story background; they may read the story silently, but more often they take turns reading it aloud with corrections of mistakes as needed by the teacher; they answer questions about the story; they occasionally receive direct instruction on some aspect of reading; and finally they receive directions for seatwork.

Although there are many questions about the effectiveness of the small-group reading lesson, its complexity has meant that the critical issues of concern to reading educators have never been satisfactorily examined. This is despite the fact that there have been great strides in research on teaching in the past decade. One problem is, as Duffy (1981, p. 113) has noted, that "reading research and teaching research have been moving forward 'out of earshot' of each other. Research on reading reflects little of what has been discovered about teaching; the research on teaching reflects few of the findings about the reading process."

The research reported in this paper was concerned with both the process of reading and teaching effectiveness. Four classes of variables were investigated, namely aspects of teaching method, setting characteristics, attributes of materials, and the reading proficiency and other personal

characteristics of the children. A noteworthy feature of the research described here is that interactions among all four classes of variables were systematically examined.

The major purpose of the present research was to try to provide a partial empirical answer to the question of under what circumstances reading instruction ought to emphasize meaning to a greater extent than word identification. The children were third graders who ranged widely in reading ability and the materials ranged widely in difficulty. Thus, the study had the potential to pinpoint the level of reading ability of children and level of readability of materials at which a meaning emphasis may get better results than a word identification emphasis.

For half of the children who participated in the study, meaning was stressed. This was done by requiring the children to predict what might happen next after reading each of a number of sentences. For instance, after reading the sentence, "The stupid child ran into the street after the ball," most children said, "He got hit by a car," "He almost got hit by a car," or sometimes, "His mother screamed at him." Inconsequential miscues during reading were ignored. If a child made a serious miscue, the teacher supplied the word and the child went right on.

With the remainder of the children, accurate word identification was emphasized instead of sentence content. Whenever a child mispronounced a word, the teacher supplied the word, and the child repeated it and finished the sentence. Then the child repeated the whole sentence. If there were still miscues, the teacher corrected them again and the child had to repeat the sentence another time. This process continued until the child gave the

sentence a fluent, word-perfect reading. Since the material was rather difficult for third graders, many poor readers had to repeat some sentences two or three times.

Over thirty factors and their two-way interactions were studied in this research. One reason so many factors were investigated is that if reading instruction is truly complex, then the questions about a meaning emphasis or a word identification emphasis cannot be answered independently of questions about other factors. While a great variety of issues of potential interest to reading educators were explored, for the sake of clarity of exposition, the discussion in the remainder of this paper is organized around five questions, including the major one about a meaning or a word identification emphasis. These questions are as follows:

1. Which is more effective in a small-group, third grade reading lesson, an emphasis on meaning or an emphasis on accurate word identification?
2. Does homogeneous grouping contribute to the effectiveness of reading lessons?
3. In a small-group reading lesson, does the child who is taking an active turn get more from the lesson than the children who are following along?
4. Which is more important, the readability or the interestingness of the materials?
5. Do the teaching methods, grouping arrangements, and materials that are most successful depend upon the reading ability and other personal characteristics of the child?

Experiment 1

Method

Subjects. The subjects were 264 third grade students from five schools in Urbana, Illinois. Four students were lost when their reading group was disrupted by a guest musician in class. An additional student was lost when he was unable to furnish a recall because he was called out of class to another part of the building. This brought the total number of subjects used in the major analysis to 259, 129 boys and 130 girls.

Only 161 of these children could be given a test measuring personality characteristics. This was due in part to attrition, since the personality test was given at a different time, but mainly to school administration concerns about possible parental objections to the test.

Materials. An independent group of 53 third grade students from Rantoul, Illinois, 31 boys and 22 girls, rated 72 sentences as to how interesting they were. From these ratings, the 36 sentences used in the study were chosen on the basis of their overall interest value and on their differential interest to boys and girls. The final list contained nine sentences that were interesting to both boys and girls (e.g., "The hungry children were in the kitchen helping mother make donuts"), nine sentences that were uninteresting to both groups (e.g., "The old shoes were put away in the back of the closet"), nine sentences that were more interesting to boys than girls (e.g., "Green blood ran out when the boy shot an arrow through the monster's head"), and nine sentences that were more interesting to girls than boys (e.g., "The crowded schoolyard was full of girls getting ready for the jump rope contest").

The sentences were written to vary widely in readability and at the same time to keep readability independent of interest. Average readability was deliberately made high for third graders, so that there would be opportunities in the word identification groups for teacher intervention, feedback, and repetition. The average readability of the sentences on the Fry scale (Fry, 1978) was 4.3 and the range was from first to seventh grade.

In order to make the position of sentences in the lesson independent of other sentence characteristics, three blocks of sentences were formed. Each block consisted of a randomly chosen one-third of the sentences from each of the four interest categories described above. The sentences within each block were then arranged in three separate random orders. Using the three block orders and the three randomization orders of sentences within each block, nine counterbalanced sentence presentation orders were formed. Nine sets of $5\frac{1}{2}$ x $8\frac{1}{2}$ booklets were assembled. Each set contained one of the presentation orders of the sentences. Each page of a booklet displayed three of the sentences.

The major dependent variable in this study was recall of the sentences from the lesson. In order to make the position of a sentence in the recall test independent of its position in the lesson, and independent of other sentence characteristics, three additional blocks of the 36 sentences were formed. Each of these three blocks contained a randomly selected third of the sentences from each of the three sentence presentation blocks. Nine counterbalanced recall orders were then developed, using the same procedure as with the presentation orders.

The standardized reading comprehension test was from the Metropolitan Achievement Test. The reading fluency test consisted of four lists of 15 unrelated words, adapted from Mason (1976), which the child read as fast as possible. A fifth list was constructed by selecting 16 difficult words from the 36 sentences used in the lesson in order to assess the effects of the lesson on speed and accuracy of word identification. A 58 question personality test was developed by adapting items from other tests used to measure introversion-extroversion (Eysenck & Eysenck, 1975), test anxiety (Hill, 1980), and internal locus of control (Crandall, 1975).

Procedure

Teachers were asked to rank order every student in their classes in terms of their judgments of reading ability. These ranked lists were then sectioned into quartiles and used to select four-child groups either heterogeneous or homogeneous in reading ability. Heterogeneous groups were created by selecting one student from each quartile while all four students in the homogeneous groups were from the same quartile. Whereas the groups were constructed based on teacher judgment, the measure of group ability used in the data analysis was a composite that gave equal weight to teacher judgment, standardized reading comprehension score, and word reading fluency score.

Initially each classroom was randomly designated for either heterogeneous or homogeneous grouping, but it was sometimes impossible to construct more than one or two complete homogeneous reading groups from each quartile. When it wasn't possible to place all students from a given quartile into homogeneous groups the remaining students were placed into

heterogeneous groups. Additional children were taken from other third grade classes in the same school if they were needed to complete a reading group.

The experimental reading instruction was conducted during the class' regular reading period. The experimental teacher, a male research assistant in his late 20's, called four students that formed a reading group to a quiet corner of the classroom where they were seated in a semicircle in front of him. The group was assigned to receive either a meaning or word identification emphasis according to a predetermined, random schedule. Initial instructions to the group were the same for both the meaning and word identification treatments, as follows:

> I have some sentences for you to read today. You'll take turns reading the sentences out loud, just like you do when you're in your reading group in class. This means that when one person is taking their turn reading out loud, the rest of you should follow along in your own booklet reading each sentence silently to yourself. Please don't say anything or make any noise when another person is reading because we want to make sure that we hear clearly every word that the person who is reading says.

The teacher then chose booklets containing one of the nine sentence presentation lists according to a predetermined, random schedule and handed one booklet to each student. Instructions for the treatment conditions were given at this point. Students to receive a meaning emphasis were told:

> When you read the sentences I want you to read each sentence so that you know what it's about. We'll be asking you some questions about the sentences later. As you read each sentence, I want you to think of a little story that the sentence could be part of, and think of what might happen next after the sentence. Then tell us what will happen next.

Instructions for the word identification treatment were:

> When you read the sentences, I want you to read each sentence carefully. We'll be asking you some questions about the sentences later. As you read each sentence, I want you to try to say every word in the sentence exactly right and clearly. If you make a mistake on a word, I'll help you, and then you should read that word again so that it's exactly right. You'll need to pay very close attention to all the words.

Both groups were given two practice sentences. All four students in the group were given the opportunity to read each of the practice sentences. In the meaning condition each student was asked to relate to the group what he or she believed would happen next, while in the word identification condition the teacher emphasized to each student the importance of reading each word carefully and saying it exactly right.

After answering any questions from the students, the teacher assigned a student to begin reading. Who would read first was based on seating position. Seating positions were labeled one through four with the student on the teacher's far left in position one and the student on the teacher's far right in position four. The starting point rotated from the seating position one to seating position four, returning to position one after every fourth reading group. Students then took turns reading the sentences orally in "round robin" fashion. A turn was defined as reading the three sentences on one page of the experimental booklet.

In the word identification treatment, the teacher promptly corrected every reading miscue that the oral reader made and then had the student repeat the word and reread the whole sentence. Even if the child spontaneously corrected an error, he or she had to reread the sentence until one perfect rendering of the sentence was achieved. When there was a long

hesitation, the teacher supplied the word, and the child finished the sentence and then read it again. Performance was praised by saying "that's right" or "good reading."

During meaning emphasis instruction, miscues were ignored unless the meaning of the sentence was disrupted, in which case the teacher supplied the word and the child went right on. When necessary, the teacher prompted for a continuation to the sentence by saying "and then what happened?" Performance was praised with "very good" rather than "that's right," in order to encourage reflection about possible meanings instead of a perception that there was a "correct" answer.

Upon completion of reading instruction, each student from the group went to one of four research assistants to receive the fluency test and the recall test. To introduce the fluency test the research assistant said:

> I have five lists of words for you to read today. I will hand you one list at a time. Your task is to read the words out loud to me as fast as you can making as few mistakes as you can. If there's a word you can't read, just say "skip it" and go on. Don't stop to try to figure out words you don't know.

The experimenter recorded all reading miscues and used a stopwatch to time the students as they read each word list. Time was recorded to tenths of a second.

Following the fluency test, the research assistant administered a test that required the student to recall the 36 sentences. One of the nine orders for sentence recall was assigned to the child based on a predetermined, random schedule. The "clue" for each sentence consisted of the subject noun and its modifier. For warm up, the student was given clues to the two practice sentences used initially. For the practice sentences

only, the clue was elaborated with a question such as, "Can you tell me what the 'fat cook' did in the sentence you read earlier?" in order to make sure the child understood the task. The child's answers were marked on a form that listed all of the sentences. The assistant crossed out the parts of a sentence that a child failed to mention and wrote down any words the child included that did not match those in the sentence.

Following recall the student was thanked for participating and sent back to class. A few weeks following reading group instruction the Metropolitan reading comprehension subtest was administered to each class. The entire class was tested as a group in its regular classroom by one of the research assistants. On another day the test measuring the various personality characteristics of the child was administered, again in the regular classroom by one of the assistants.

A simple, objective method of scoring sentence recall was employed in the main data analysis. The number of content words that a child reproduced from a sentence was counted. This count included not only words recalled verbatim but also synonymous words and phrases. Then, since the sentences varied in length, the number of content words recalled was divided by the number of content words in the sentence (not including the subject noun phrase which comprised the clue). The child's recall of the sentences was expressed as a percentage. A subsidiary analysis used a more lenient scoring system in which a child got full credit for recalling the gist of any part of a sentence.

Results and Discussion

Overview of Analysis

Table 1 includes the mean, standard deviation, and range for each variable except derivative variables such as the fluency of the oral reader, the interest value of the prior sentence, the quadratics of continuous variables, and interactions. Table 2 summarizes the analysis involving the principal dependent variable, percentage of sentence recall. Factors are listed in order of entry into the equation. This table contains just the "reduced" model, which was compiled by deleting all nonsignificant interactions and rerunning the program. Also deleted from the reduced model were nonsignificant main effects, provided the factors were not involved in any significant interactions. Altogether about 200 main effects and interactions were examined. The examined terms included all two-way interactions except those among child characteristics (and those involving one other factor). None of the possible 4060 three-way interactions and none of the interactions involving four or more factors was included. There was no theory to guide the selection of higher-order terms. Aimless exploration would have vitiated the power of the study and yielded nothing except a large computer bill.

Insert Tables 1 and 2 about here.

No computer program was available that could handle the entire analysis in one pass. It was necessary to run a truncated form of the complete model that included 140 variables, delete variables representing nonsignificant interactions and quadratic terms, add the remaining 60 variables, and run the program again. The program used was BMD-P2R.

Table 1

Description of Measures

Variable	Measurement Scale/Method	Meaning of High, Positive Score	Mean	SD	Range	
					Smallest Value	Largest Value
Sentence Recall	Percentage	Good Recall	28.51	14.30	2	69
Instructional Time	Minutes	Long	13.09	3.79	6	22
Word Errors	Frequency	Many	3.61	4.73	0	16
Word Reading Time	Seconds	Slow	26.18	22.14	6.4	134.3
Reading Comprehension	National Stanine	High	5.04	2.05	1	9
Reading Fluency	Local Stanine	Fluent	5.03	1.93	-.75	9.02
Sex	Contrast Coded	Boy	.00	1.00	-1	1
Ethnicity	Contrast Coded	White	.45	.89	-1	1
Introversion-Extroversion	Self-Report	Extroverted	12.03	2.80	6	20
Anxiety	Self-Report	Not anxious	8.54	3.57	1	14
Responsibility Accepted	Self-Report	High	8.37	1.90	3	12
Blame Accepted	Self-Report	High	5.47	2.32	0	11

Table 1 (Cont'd)

Variable	Measurement Scale/Method	Meaning of High, Positive Score	Mean	SD	Range	
					Smallest Value	Largest Value
Attention Span	Percentile Rank	Long	51	29	2	98
Meaning vs. Word Identification	Contrast Coded	Meaning Emphasis	-.03	1.00	-1	1
Group Ability	Local Stanine	High	5.02	1.27	2.1	7.4
Group Homogeneity	Standard Deviation	Heterogeneous	1.18	1.12	.19	2.25
Active Turn?	Contrast Coded	Yes	-.50	.87	-1	1
Cue-Response Association	Adult Rating	Highly Associated	4.20	1.27	1.14	6.93
Readability	Grade Equivalents	Difficult	4.28	1.68	1.0	7.0
Position in Lesson	Serial Position	End	18.50	10.39	1	36
Position in Test	Serial Position	End	18.50	10.39	1	36
Total Interest	Peer Rating	Interesting	1.52	.88	.27	2.82
Differential Interest	Peer Rating	Girl Oriented	.06	.46	-.61	1.10

Table 2

Summary of Regression Analysis for Experiment 1

Variable	Regression Coefficients		% Variance	F
	Between	Total		
Between Subjects				
Reading Comprehension	4.22	2.49	10.97	35.56
Reading Fluency	- .87	- .35	1.69	5.48
Sex	1.92	1.99	1.74	5.63
Meaning vs. Word Identification	-8.48	2.80	3.81	12.34
Group Ability	-2.21	- .73	2.66	8.63
Meaning vs. Word Identification x Group Ability	2.26	--	3.80	12.31
Constant/Residual	22.75		75.33	
Within Subjects				
Cue-Response Association		5.42	5.11	498.80
Readability		.01	.07	7.09
Position in Lesson		- .07	.01	1.13
Position in Test		- .15	.21	20.27
Total Interest		7.87	1.65	161.01
(Total Interest)^2		-2.63	.09	8.90
Differential Interest		11.48	.00	.07
Active Turn		.52	.52	50.67
Reading Fluency of Active Turn Taker		-1.65	.17	16.89
Position in Lesson x Active Turn		.12	.10	10.13

Table 2 (cont.)

Variable	Regression Coefficients		% Variance	F
	Between	Total		
Position in Lesson x Total Interest		.14	.09	8.90
Meaning vs. Word Identification x Readability		-1.20	.23	22.63
Meaning vs. Word Identification x Reading Fluency of Active Turn Taker		2.26	.28	27.02
Meaning vs. Word Identification x Differential Interest		1.03	.07	6.98
Group Ability x Differential Interest		-2.35	.18	18.02
Sex x Differential Interest		-3.91	.24	23.19
Reading Comprehension x Total Interest		.65	.10	9.46
Constant/Residual		- .10	90.88	

Note: The degrees of freedom for between-subject factors are 1 and 244; the degrees of freedom for within-subject factors are 1 and 8865. An F of 8.18 is significant at the .01 level, two-tailed.

The program was run in strictly hierarchical order. In the portion of the analysis in which the total variance was partitioned, the variables were entered in blocks as follows: within-subject factors, quadratics of selected within-subject factors, interactions among within-subject factors, between-subject factors, and, finally, interactions between within-subject and between-subject factors. Within blocks, order of entry was based on logic or theory. In cases in which no a priori grounds were discernible, variables were entered stepwise in order of variance accounted for. The between-subject factors were evaluated in a separate analysis; they were included in the total analysis in order to get proper estimates of interactions of within-subject and between-subject factors and appropriate beta weights.

The F ratio for each factor was constructed off-line based on the increment in R^2 at the point of entry. Naturally, between-subject factors were evaluated in terms of between-subject variance and within-subject factors in terms of within-subject variance. In each case, the error variance was the final residual after all variables had been entered into the equation.

The unit of analysis was the individual subject's performance on individual sentences. There were 36 sentences and in the main analysis there were 259 subjects; hence, there were 9324 observations. The degrees of freedom for within-subjects tests of significance in the main analysis was $N(K-1)-V = 8865$, where N is the number of subjects, K is the number of sentences, and V, which equalled 200, is the number of variables tested for significance. In the subsidiary analysis involving the noncognitive

characteristics of the children there were 161 subjects and 5445 degrees of freedom.

The nominal alpha level for each test of significance was .01. Significance was probably overestimated somewhat since the assumption of equal sentence by sentence intercorrelations was violated. Because about 200 tests were performed, the expected number of false alarms was approximately two. In order to avoid clutter, most effects will be referred to in the text simply as "significant" or "nonsignificant." The supporting detail for significant effects can be found in Table 2.

The percentage of variance explained by within-subject factors may appear to be rather small. The value seems small because we did not aggregate across subjects to get a single average result for each sentence, which has been the standard practice in readability research (e.g., Bormuth, 1966) and psycholinguistic research (e.g., Just & Carpenter, 1980). Aggregation of results causes huge increases in proportion of variance that is apparently explained. However, this is a practice that requires the untenable assumption that there are no systematic individual differences among subjects, in particular that characteristics of subjects and attributes of tasks, settings, and materials do not interact. "Accounting for variance" is not the point of educational and psycholinguistic research. Rather, the goal is to understand human comprehension, learning, and remembering as these processes may be conditioned by task, materials, and setting.

The present study used a longer list of sentences than is customary in experiments with children. This was done in order to appraise what would happen when the novelty of the experiment had worn off somewhat. An

undesirable side effect of a long list was that level of sentence recall was low, only 28.5%. An experiment is like a test in the sense that its discriminating power is greatest when the average level of performance is around 50%. When the level deviates substantially from 50% the relationships among variables can be distorted and spurious interactions can arise because of "ceiling" or, in this case, "floor" effects.

Another problem when level of performance is low is that the absolute size of effects is constrained. This is not inherently a problem, except that it may lead the unwary to conclude that though an effect is significant its size is not large enough to be practical. The absolute size of an effect is a poor guide to decision making. A better basis would be provided by getting a sense of how the experiment is "calibrated." This can be done by examining the size of the effect produced by a familiar benchmark variable. In the present study the benchmark variable might be performance on the standardized reading comprehension test. On the average, before any other factors were considered, each stanine increase in reading comprehension score resulted in a 2.15% increase in sentence recall. Considering the full range of performance on the reading comprehension test, on the average a child who scored at the 9th stanine recalled 19.4% more sentences than a child who scored at the 1st stanine.

The remainder of this section is organized around the five questions raised in the introduction. The answers to these questions are not independent. This means that it will be necessary to discuss some results several times in different contexts.

Meaning Emphasis Versus Word Identification Emphasis

The first question is, which is more effective in a reading lesson, an emphasis on meaning or an emphasis on accurate word identification? The answer from this study is that an emphasis on meaning significantly promoted mastery of the lesson, as evidenced by the children's recall. On the average children in groups that received a meaning emphasis recalled 5.7% more sentences than children in groups that received a word identification emphasis. However, this finding has to be qualified because the teaching emphasis interacted significantly with three other factors.

*The teaching emphasis interacted with the readability of the sentences. Figure 1 shows that the easier the sentences the greater the advantage of a meaning emphasis.

*The teaching emphasis interacted with average reading ability of the group. This interaction will be discussed in the next section on grouping.

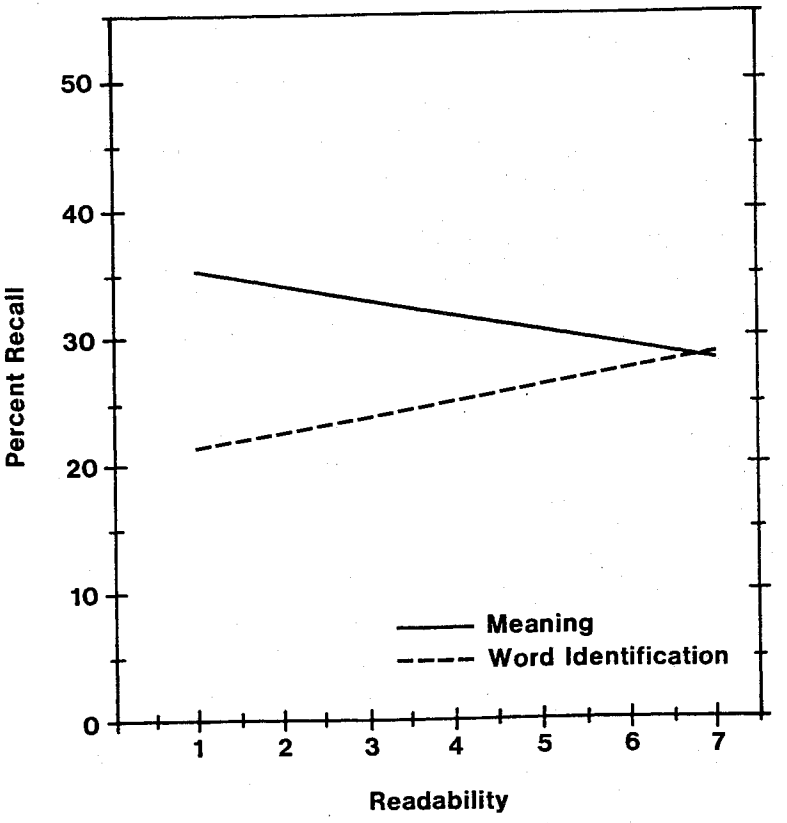
*The teaching emphasis interacted with the fluency of the child whose turn it was to read aloud. This interaction will be discussed in the section on turn taking.

Insert Figure 1 about here.

It is most interesting that the expected positive relationship between readability and performance was actually reversed in the groups in which accurate word identification was stressed (see Figure 1). This is readily understandable when you consider the instructional procedure in these groups. To reiterate, the experimental teacher corrected each miscue and then the child read the sentence again. If there were still miscues the

Figure 1

Percentage recalled as a function of teaching emphasis and readability.



whole procedure was repeated again until the child was able to give the sentence a perfect reading. Thus, a difficult sentence provided a substantial opportunity for feedback and rehearsal. In contrast, an easy, flawlessly read sentence whizzed by with little opportunity for learning.

An important finding is that there was no interaction between the child's reading ability and the teaching emphasis, $F < 1.0$. On its face this finding points to the conclusion that a meaning emphasis gives better results than a word identification emphasis with third graders at every level of reading proficiency. However, we are obliged to report that this conclusion is not entirely data driven. It hinges on the policy we adopted for doing the data analysis. This policy caused us to enter the group ability by teaching emphasis interaction into the equation before entering the individual children's reading comprehension ability by teaching emphasis interaction. An alternate analysis showed that the interaction with the measure of individual ability is nearly significant, $F(1,244) = 7.22$, $.01 < p < .02$, % Var = 2.23, $B = 1.07$ when it is entered first.

It should be stressed that the facts do favor giving priority to the group ability by teaching emphasis interaction. When this interaction is entered into the analysis first, the amount of variance explained by the individual ability by teaching emphasis interaction is nil, as we have just explained. On the other hand, when the individual ability by treatment interaction is entered first, the group ability by treatment interaction is almost significant, $F(1,244) = 6.08$, $.01 < p < .05$, % Var = 1.85, $B = 2.54$. These facts suggest that everything in the individual interaction term is also represented in the group interaction term, but that the group term may include something else of importance in addition.

Whether there is really an interaction between the level of reading proficiency of the child and the type of teaching emphasis is an issue with important educational policy implications. It would be unacceptable to let a conclusion about an issue of such importance depend upon the approach used to perform an analysis or upon what could be undependably fine nuances in a set of data. Therefore, we took the position that the possible interaction of individual reading comprehension ability with treatment had to be given serious attention.

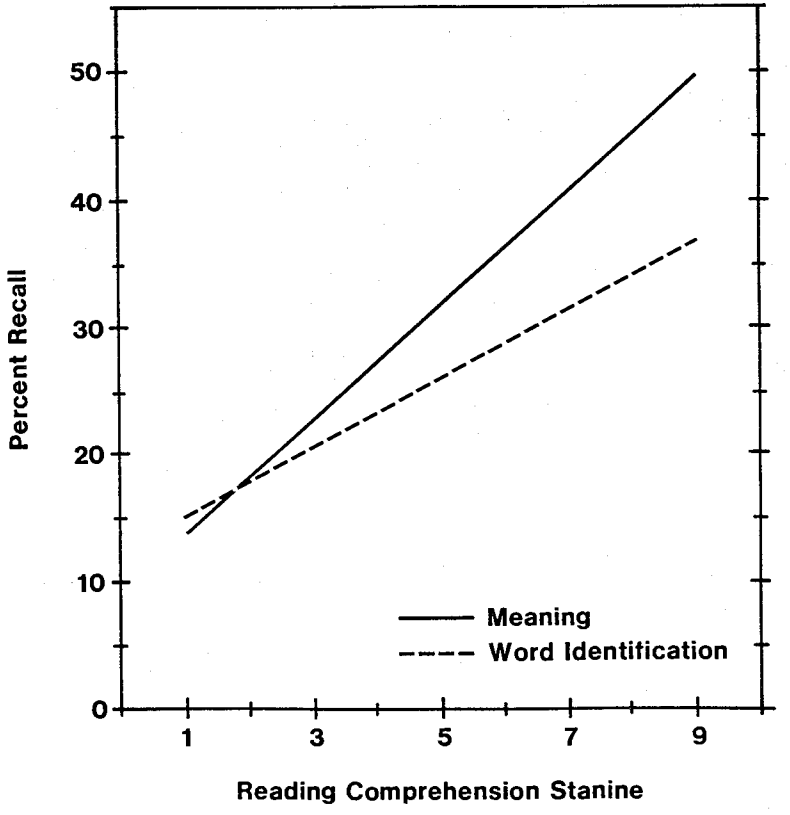
A graph of the possible interaction is presented in Figure 2. If the interaction is accepted as trustworthy, then the higher the ability of the child the greater the advantage of an emphasis on meaning. Conversely, when the ability of the child is low, there is little difference between treatments, or even a slight advantage for a word identification emphasis.

Insert Figure 2 about here.

It must be cautioned that the possible small benefit of a word identification emphasis with very slow children shown in Figure 2 may be due to an artifact. The level of learning in this study was low because the list of sentences was long and average difficulty was high. Thus, as we already explained, there may have been a performance "floor." In other words, the performance of poor readers may have been so low already that there was no room on the downside for even poorer performance to show itself in the word identification condition. Experiment 2 was designed to see whether there is a real benefit from a word identification emphasis for children with low ability, or whether the possible slight advantage in this study was an artifact.

Figure 2

Percentage recalled as a function of teaching emphasis and individual reading comprehension ability.



The principal reason for the difficulty in determining whether there was really an interaction between the individual child's reading ability and the teaching emphasis was the correlation (r = .56) between individual ability and average group ability. A simple strategem was used in Experiment 2 to unconfound the effects: The children were instructed individually, so the issue of group ability did not arise.

Grouping for Instruction

The question that we shall attempt to answer in this section is whether grouping children according to ability increases the effectiveness of reading lessons. Allow us to emphasize again that in this study we have defined effectiveness in terms of degree of mastery of a single day's lesson, as indicated by the percentage of the sentences from the lesson that the children were able to recall.

The measure of homogeneity was the standard deviation of the composite reading scores of the children in a group. This variable had absolutely no effect and did not enter into any interactions. However, the mean level of reading ability of the groups did have a substantial effect and was involved in two significant interactions:

*As reported in the preceding section, group ability interacted with the teaching emphasis. Figure 3 indicates a strong negative relationship between group ability and sentence recall in groups receiving a word identification emphasis, but no relationship in groups receiving a meaning emphasis.

*Group ability interacted with the relative interest of sentences to boys and girls. This interaction is discussed in the section on materials.

Insert Figure 3 about here.

To dispel some of the mystery about the negative relationship shown in Figure 3 between group ability and sentence recall in the word identification groups, let us emphasize again that there was a positive relationship between the individual child's reading ability and recall; however, regardless of a child's own ability he or she recalled more of the sentences when the average group ability was low. The principal reason this happened is that the lower the ability of a group the longer was the instructional time. In groups receiving a word identification emphasis, but not those getting a meaning emphasis, instructional time was strongly related to lesson mastery. These facts are embodied in the causal model diagrammed in Figure 4A. The arrows mark causal paths. An arrow between two boxes indicates that the first factor is hypothesized to be the cause, or part of the cause, of the second factor. Paths from unmeasured causes are not included. The structural equations for this model, written in unstandardized form, are as follows:

$$\underline{x}_3 = \underline{a}_3 + \underline{p}_{13}\underline{x}_1 + \underline{p}_{23}\underline{x}_2$$

$$\underline{x}_4 = \underline{a}_4 + \underline{p}_{34}\underline{x}_3$$

$$\underline{x}_5 = \underline{a}_5 + \underline{p}_{15}\underline{x}_1 + \underline{p}_{45}\underline{x}_4$$

where $\underline{x}_1$, $\underline{x}_2$, and $\underline{x}_3$ are expressed in stanines, $\underline{x}_4$ is expressed in minutes and $\underline{x}_5$ is expressed in percentage of recall.

Insert Figure 4A about here.

It is clear that the model gives a good account of the dynamics in the groups that had a word identification emphasis. Each stanine decrease in

Figure 3

Percentage recalled as a function of teaching emphasis and group ability.

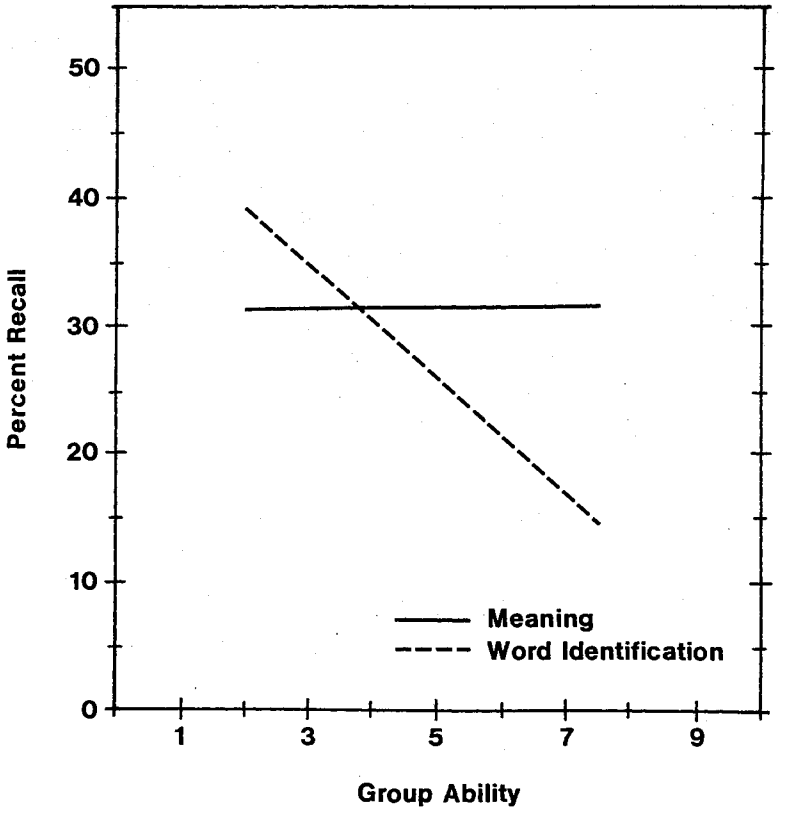
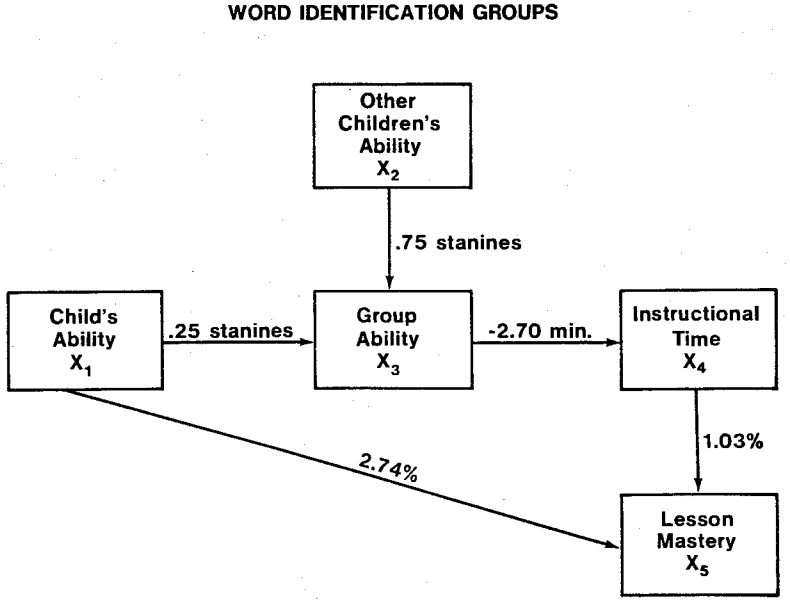


Figure 4A

Partial causal model of relations between reading ability, instructional time, and lesson mastery.



group ability caused 2.7 minutes increase in instructional time, and each minute increase in instructional time caused a 1.03% increase in sentence recall. Thus, one stanine change in group ability led via an effect on time to a 2.7 x 1.03 = 2.78% change in lesson mastery.

It is equally clear from Figure 4B that the model does not capture the dynamics of meaning emphasis groups. Whereas decreases in group ability led to increases in instructional time, there was only a very slight connection between instructional time and sentence recall. The net effect was only a 1.78 x .07 = .12% change in lesson mastery when there was a one stanine change in group ability.

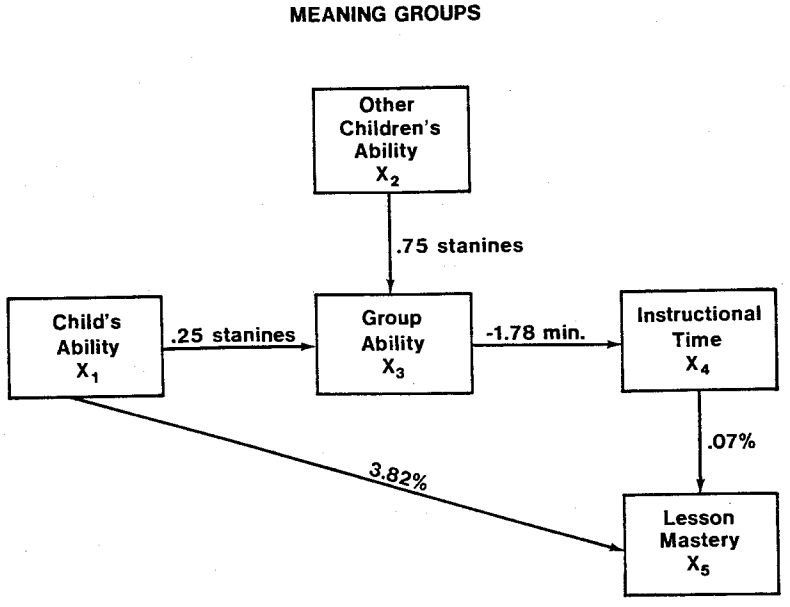
Insert Figure 4B about here.

The interesting conclusion that is clarified by the causal model is that lessons with a word identification emphasis depend heavily on time for their effectiveness. Instructional time makes little difference in meaning emphasis lessons; here it is the quality of the child's encounter with the lesson material that is important, not the quantity.

Based on other research we may be reasonably confident that there is one aspect of these conclusions which will <u>not</u> generalize to natural classroom settings. In actual classrooms there is typically a <u>positive</u> relationship between functional instructional time, or "opportunity to learn," and group ability (Hiebert, 1982). In the special circumstances that prevailed in the present study the experimental teacher allocated more time to make up for the slower pace in the low groups. Teacher effectiveness research shows that student achievement is a function of the pace (that is, the rate at which the curriculum is covered), allocated time, and the proportion of

Figure 4B

Partial causal model of relations between reading ability, instructional time, and lesson mastery.



allocated time in which the student is actually engaged (cf. Berliner & Rosenshine, 1977; Leinhardt, Zigmond, & Cooley, 1981). Furthermore, classroom research suggests that the pace is slower in low ability groups (Beckerman & Good, 1981) and that, whereas allocated time is about the same in high and low groups, usable instructional time and engagement rates are considerably lower in low groups (cf. Allington, 1980).

Thus, there is reason to doubt the overall generality of the results encapsulated in the model diagrammed in Figure 4A because we made sure all groups covered all of the material. However, there is one aspect of the conclusions that deserves close attention from instructional researchers. We know of no reason to doubt the generality of the finding that the success of meaning emphasis instruction depended very little on instructional time. This method succeeds because of the quality of the children's encounter with lesson materials. Could it be that the reason that instructional time has received such prominence in recent years is that the methods employed by classroom researchers are capable of identifying instruction whose degree of success depends upon time but incapable of distinguishing methods whose success depends upon qualitative characteristics?

Turntaking During Instruction

The principal question that we attempted to answer is, does the active participant (the child who is called on to read aloud and respond to the teacher's instruction) get more from the lesson than the other participants (the remaining children in the group who are reading silently and listening)? The answer based on this study is "yes." Children achieved

significantly greater lesson mastery when they were taking a turn than when they were following along. The mean advantage was 6.1%. To our knowledge, this study provides the first direct evidence of the benefits of playing an active role, even though educators have always believed that active engagement is important. However, further research will be required to determine whether active engagement was really the operative factor. Since any one child's turns encompassed only a quarter of the sentences, these sentences were, therefore, more distinctive than the remaining three quarters. Thus, distinctiveness could have been the operative factor instead of active engagement.[2]

There was one significant interaction involving role:

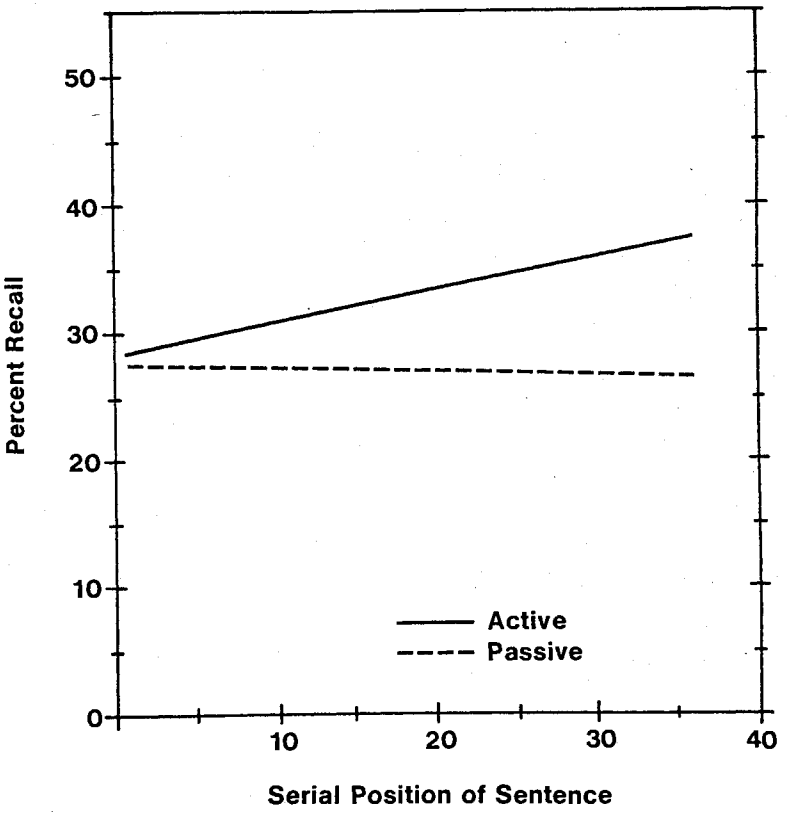
*Whether a child was taking a turn interacted with the position of the sentence in the lesson. Figure 5 shows that when the child was the active participant performance improved through the course of the lesson, whereas performance of the other participants remained about the same.

Insert Figure 5 about here.

One reason we used the predictable, "round robin" turntaking procedure was to make it obvious when a turn would begin so that we could determine whether there was psychological preparation for turntaking that interfered with the processing of material then being covered. Specifically, we looked at recall of sentences that immediately preceded active turns. No effect was observed, $F < 1.0$. Performance was indistinguishable from other sentences encountered when the child was playing a passive role. It should be noted that the physical layout of the materials made it difficult for a

Figure 5

Percentage recalled as a function of turn taking and position of sentence in lesson.



child to actually practice the sentences he or she was going to have to read aloud, because they were always printed on the next page.

It could be that taking an active turn engages attention and that, once engaged, attention persists for a period after the turn has ended. This possibility was evaluated by examining performance on sentences that immediately followed active turns. The finding was negative, $\underline{F} = 1.37$. Performance was no different from performance on other sentences encountered when the child was playing a passive role.

The active participant's reading fluency was a salient characteristic that figured in two significant effects. The measure of fluency, it will be recalled, was speed of reading four lists of 15 words. The fluency of the oral reader had a significant negative effect on lesson mastery. On the average, there was a .8% decline in sentence recall for each stanine increase in the fluency of the oral reader. The oral reader's fluency was involved in a significant interaction:

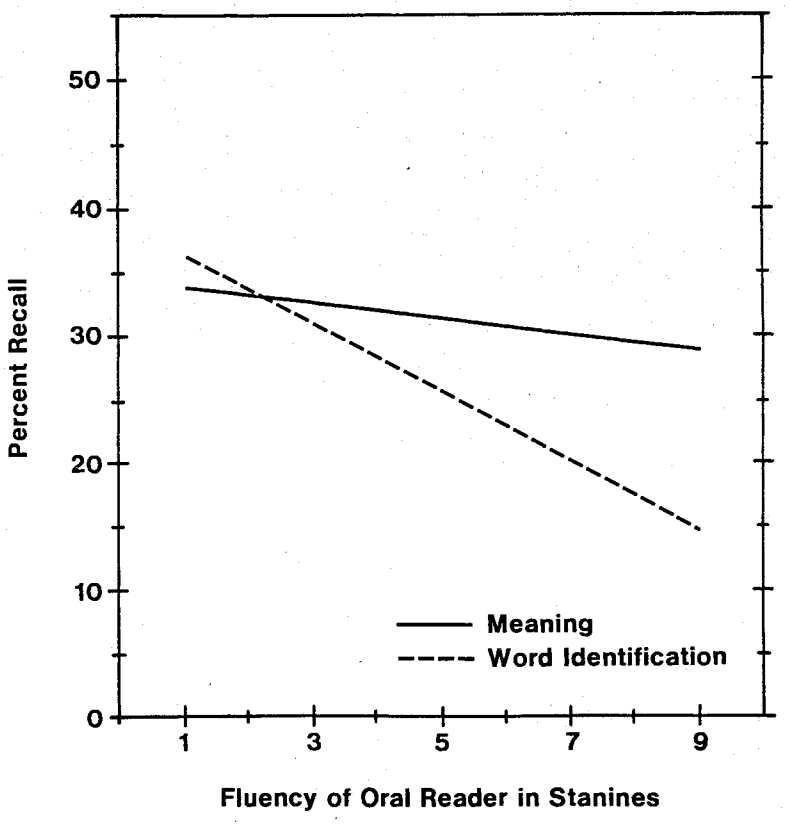
*The fluency of the oral reader of the moment interacted with the teaching emphasis. Figure 6 indicates that lesson mastery was a sharply decreasing function of the oral reader's fluency when there was a stress on word identification, but that the oral reader's fluency made less difference when there was a stress on meaning.

Insert Figure 6 about here.

The negative effect of the fluency of the oral reader in the word identification emphasis groups can be understood in terms of the causal model introduced earlier. Only now we can refine the model. It is the fluency of the oral reader that is critical, not other aspects of ability.

Figure 6

Percentage recalled as a function of teaching emphasis and the fluency of the oral reader.



Of course, there is a moderately high correlation between reading fluency and reading comprehension. In this study the correlation was .62. Often when variables are intercorrelated it is impossible to identify the one which has the causal force; however, in this case a clear determination is possible. The critical fact is this: The reading comprehension of the oral reader, when entered into the equation instead of the fluency of the oral reader, does not have a significant effect ($F = 1.17$), despite the correlation between the two measures. This proves that the fluency of the oral reader is the operative aspect of his or her reading ability.

It is entirely reasonable that fluency, rather than comprehension, is the important aspect of children's ability when they are reading aloud in a reading group, at least when errorless performance is the goal. The teacher and the other children can readily distinguish a stumbling, halting rendition of a sentence from a smooth performance. Comprehension, on the other hand, is an internal process with fewer manifestations perceptable to listeners.

Figure 7 presents an augmented causal model relating two aspects of reading ability, the fluency of the oral reader, instructional time, and lesson mastery. Notice that this model differs from the one displayed in Figure 4A in two respects. The first difference is the separation of reading ability into two facets. The preceding paragraphs have presented the reasoning and evidence for the claim that fluency, not comprehension ability, lies on the indirect path to lesson mastery mediated by instructional time. However, it is comprehension, not fluency, that lies on the direct path. The evidence for this is that when the fluency measure is entered in the equation after the comprehension measure it has a nearly

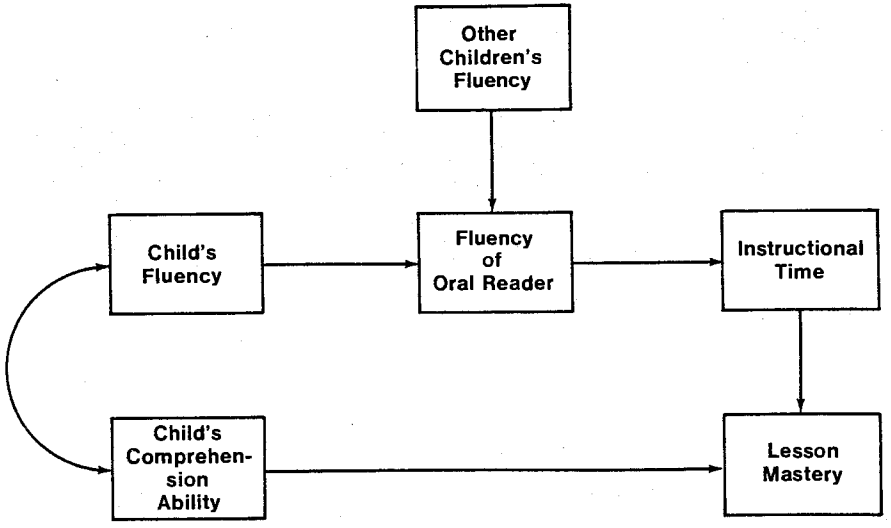
significant negative relationship with lesson mastery. To state this in other language, once the common component of the two tests has been factored out, the remaining trait measured by the fluency test depresses performance.

Insert Figure 7 about here.

The second difference between the model diagrammed in Figure 4A and the augmented model is that fluency of the oral reader, rather than the group's ability, is placed on the path mediated by instructional time. To the extent that the fluency of the children varies, so too will the instructional time on the segments of the lesson during which the children are taking their turns. Regrettably, we recorded only total lesson time, rather than time on the segments encompassing the turns; thus, it is impossible to provide quantitative estimates of the parameters of the model set forth in Figure 7. Another problem is that the fluency of the oral reader is confounded with group ability. Either the oral reader fluency by treatment interaction or the group ability by treatment interaction, but not both, will account for significant variance. This means that no strictly data-driven interpretation is possible. However, in this instance, it is our judgment that logic dictates that priority be given to the interaction involving the oral reader's fluency. We believe, therefore, that the model diagrammed in Figure 7 gives the best picture of the dynamics of the word identification emphasis groups under the conditions that prevailed in this study. Of course, the model does not fit the meaning emphasis groups.

Figure 7

Augmented causal model relating reading abilities, instructional time, and lesson mastery.



Materials

The major question we asked about materials was, which is more important, readability or interest? This study yields a resounding answer: "Interest." Interest accounted for 33.8 times as much variance in sentence recall as readability. This value does not include the interactive and derivative effects of the two factors. When these are counted, interest explains four times as much variance as readability.

The flacid effects of readability cannot be explained away because of restriction of range. In fact, the sentences ranged in readability from the first grade to the seventh grade level. Thus, there is no escape from the conclusion that interest is much more important than readability when the criterion is recall of information read and, of course, when there is a teacher available to help the children with the hard words.

All of the multivalued variables included in the study were checked to see if they had a curvilinear relationship with sentence recall. Only interest did, as is evidenced by the fact that the quadratic of rated interest was significant. As sentences move from uninteresting to interesting there is a sharp rise in performance, but the increase is more gradual as the sentences move from interesting to very interesting. That the relationship is curvilinear is not especially noteworthy, since the interest scale is arbitrary.

The rated interest of sentences interacted with two other factors.

*Interest interacted with the position of the sentence in the lesson. Figure 8 shows that interest was more important for sentences appearing late in the lesson than it was for sentences early in the lesson.

*Interest interacted with reading comprehension ability. Figure 9 indicates that there was a greater difference in the performance of high and low ability children on interesting sentences than on uninteresting sentences.

Insert Figures 8 and 9 about here.

It should be noted in passing that Figures 8 and 9 were constructed by selecting two extreme values of continuous variables. However, in each case the calculations to determine the best-fitting functions were performed on the entire array of 9,324 observations.

We interpret the interaction of position and interest in the following fashion. An experiment is a novelty, at least at the beginning. When the session starts the children are giving some of their attention to adapting to the unfamiliar teacher and the unfamiliar procedures. As the novelty wears off and they become more comfortable with the situation, factors of intrinsic importance such as the interest of the materials are likely to take hold. We believe the same explanation can also account for the increasing importance of active participation from the beginning to the end of the lesson (see Figure 5). If this interpretation is correct, it is reasonable to infer that interest and active participation would be even more important under the daily routine of school than in a brief experiment.

If one were to take the interaction between interest and comprehension ability (Figure 9) at face value, it would seem to indicate that bright children are more sensitive to the interestingness of reading material than less able children. This conclusion does not agree with common observation. It seems possible, instead, that the interaction appeared in this study because of a performance floor which led to a distorted estimate of the

Figure 8

Percentage recalled as a function of interest and position of the sentence in the lesson.

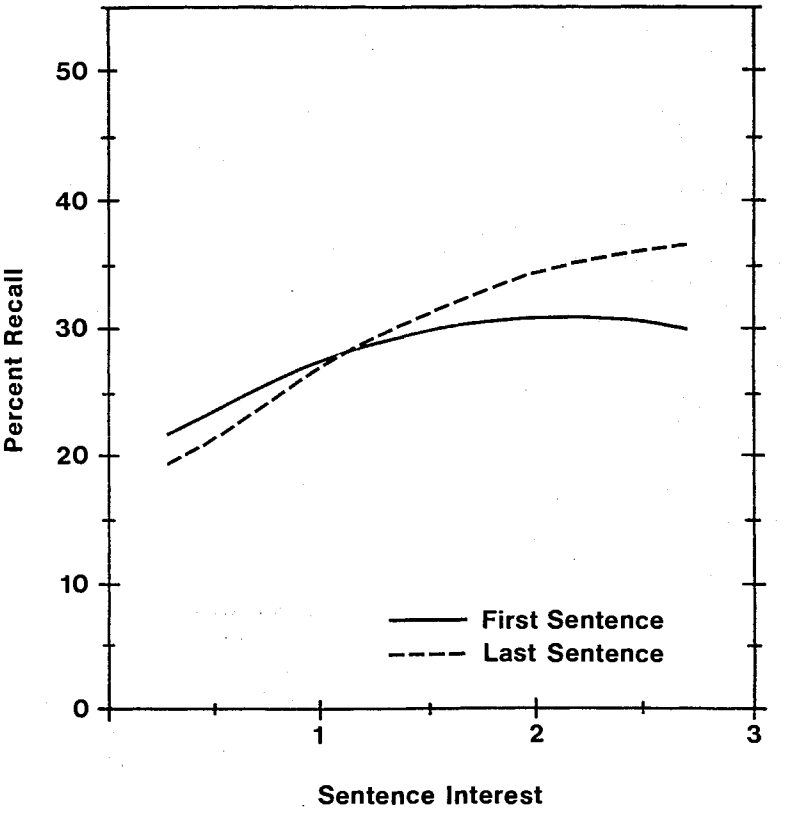
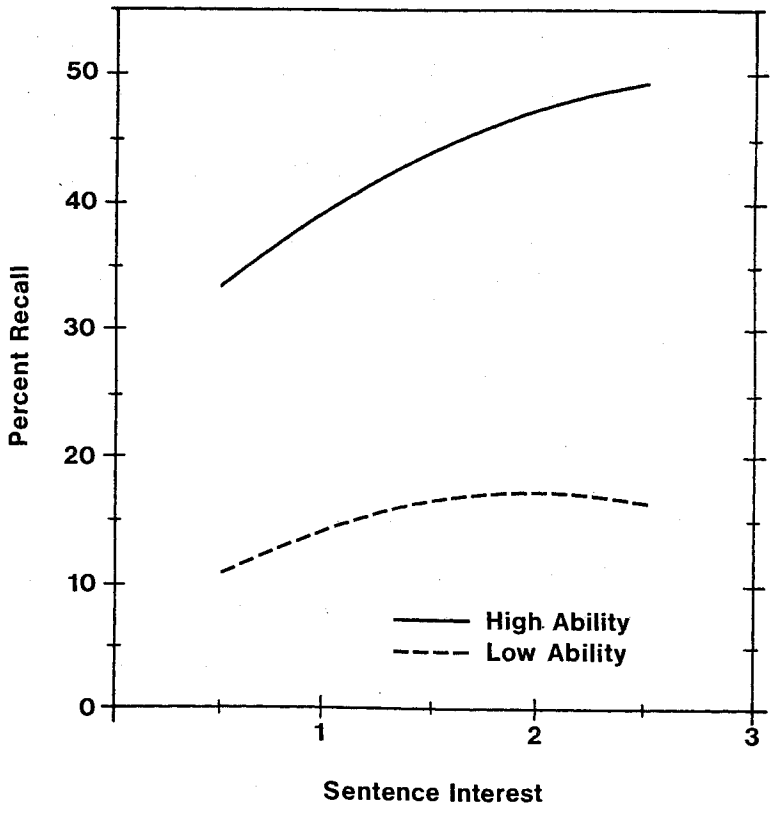


Figure 9

Percentage recalled as a function of interest and reading comprehension.



relative performance of low ability children. The relationship between interest and ability was explored further in Experiment 2.

It is a common practice to sprinkle colorful vignettes into children's social studies and science texts (Pearson, Gallagher, Goudvis, & Johnston, 1981). Often these vignettes have no conceptual relationship with the rest of the text. Apparently the theory is that interesting asides will attract attention and, once attracted, attention will be maintained for awhile, leading to better learning of the surrounding, less interesting material. We tested this theory by determining the influence of the interest of each sentence on the recall of the sentence which followed it in the lesson. There was absolutely no effect, $F < 1.0$. Also investigated was the possible influence of the interest of a sentence on the recall of the sentence which immediately preceded it in the lesson. Again, there was no effect, $F = 2.27$. Hence, there was no support for the theory that an interesting but unrelated piece of information will improve the learning of surrounding information. Evidently children switch attention on and off very rapidly. In fact, in related research (Anderson, 1982) we have been unable to establish that amount of attention is even on the causal path between interest and learning.

It is a reasonable conjecture that intrinsically interesting sentences increase learning by affecting the quality of children's processing. We have already uncovered evidence, discussed at length in preceding sections, that a teaching emphasis on meaning affects learning because of the quality rather than the quantity of processing. Hence, it is a plausible hypothesis that the influence of the two factors is mediated by the same underlying, qualitative process. An implication of this hypothesis is that the effects

of a meaning emphasis and interest will not be additive, since if the process has already been evoked by an interesting sentence having a meaning emphasis will be at least partially redundant. The results show, however, that the effects are additive, as evidenced by the failure of even a hint of an interaction to appear, $F < 1.0$. The conclusion is that the effects of a meaning emphasis and interest seem to be produced by different underlying processes.

Up to this point we have been discussing the total rated interest of sentences to third graders. We also investigated a derivative factor, differential interest to boys and girls, where a positive score means that the sentence was more interesting to girls and a negative score means that it was more interesting to boys. Of course, the absolute magnitude of the number indicates the size of the differential interest. Differential interest did not have a significant main effect, but it did enter into three significant interactions as follows:

*The differential interest of sentences to boys and girls interacted with the sex of the child. Figure 10 shows that performance was best when the orientation of the sentence matched the child's sex.

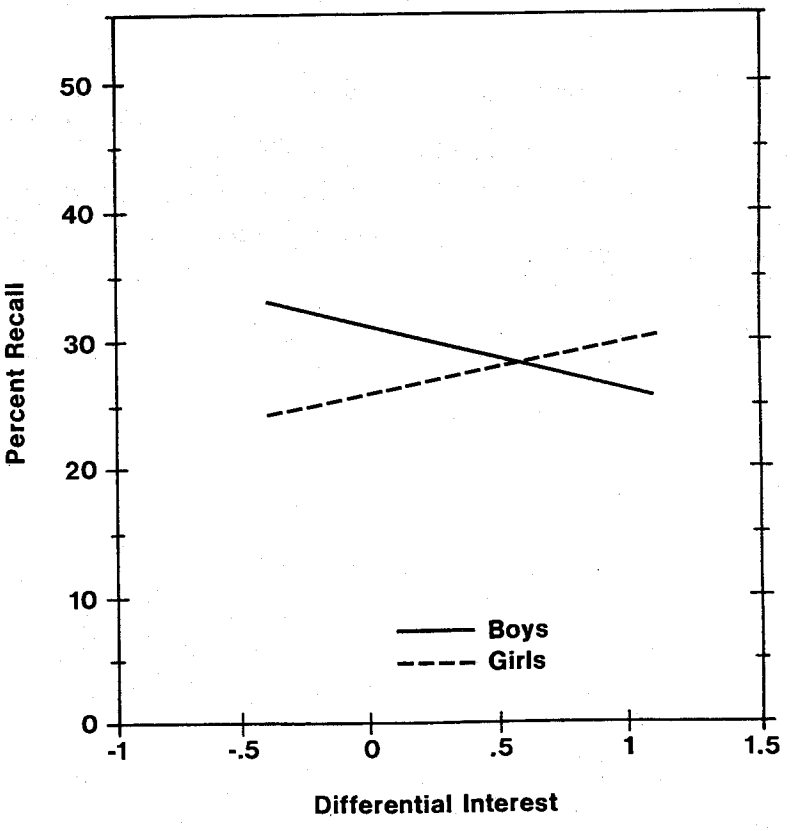
Insert Figure 10 about here.

---

*The differential interest of sentences to boys and girls interacted with the teaching emphasis. Relatively speaking, a meaning emphasis was better with girl-oriented sentences while a word identification emphasis was better with boy-oriented sentences.

Figure 10

Percentage recalled as a function of sex and differential interest.



*The differential interest of sentences to boys and girls interacted with group reading ability. Performance on girl-oriented sentences was relatively better in low ability groups than high ability groups whereas group ability made little difference on boy-oriented sentences.

The interaction between differential interest and the sex of the child is not surprising, but we were gratified to see it appear because it replicated the findings of Asher and his associates (see Asher, 1980) who used entirely different methods. There was also a nearly significant interaction between total interest and sex, $F(1,8865) = 5.30$, $.01 > p < .05$, % Var = .06, $B = 1.09$, suggesting that boys may be more sensitive to interest than girls. This, too, is consistent with Asher's findings.

We have no idea why differential interest interacted with the teaching emphasis and group ability. Perhaps these are cases whose explanation would have become clearer if the sex of the child had been figured in; however, we were unwilling to open the Pandora's Box of higher order interactions.

The findings regarding readability have already been reported. To review, readability did not have a significant main effect, but it did interact with the teaching emphasis (see Figure 1).

In analyses that we will not report in detail, readability was represented in terms of two factors, the length of each sentence in syllables and the frequency of usage of the least frequent word in each sentence. These factors accounted for somewhat more variance than Fry readability level, but there was no important change in the conclusions. Hence, Fry readability level was reported since it is familiar to a wider audience.

There were two other significant aspects of the materials that have not been discussed yet. One was the rated degree of association between the subject noun phrase of a sentence, which served as the cue in the test, and the rest of the sentence. Obviously when the cue words makes it easy to think of the rest of the sentence recall will be higher. Indeed, the degree of connection between the cues and the sentences turned out to be the most potent factor in the experiment. However, this fact is of no general pedagogical or psychological interest. The factor was included in the analysis just in case it happened to be confounded with other factors such as readability or interest, in which case serious misinterpretations would have been invited. Connectivity ratings were roughly orthogonal to other factors and the method of hierarchical analysis adjusted for the slight biases that were present. Interaction terms involving connectivity were not computed because they were of no theoretical or practical interest.

The final significant materials factor was the position of a sentence in the test. For each additional sentence there was a .14% decline in sentence recall or a total decline of 5.2% from the beginning to the end of the test. We attribute the decline to fatigue. The position of sentences in the test did not interact with any other factors.

Child Characteristics

The general question raised in the introduction is whether the success of teaching methods, grouping arrangements, and materials depend upon the reading ability and other characteristics of the child. Most of the evidence bearing on this question has already been introduced and will merely be reviewed here.

There are just two child characteristics that had significant main effects--standardized reading comprehension test score, and the word reading fluency of the child taking an active turn. It is well-known that ethnicity is related to performance on school tasks, and this study was no exception (ethnicity correlated .20 with sentence recall). However, we chose to enter ethnicity into the equation after the child's reading comprehension and reading fluency scores. In effect the question asked was, after taking account of a child's reading proficiency, can additional information be gleaned by looking at the child's skin color or listening to his or her dialect? The answer from this study was "no." Ethnicity did not produce a significant main effect and was not involved in any significant interactions. The other child characteristics did not yield significant main effects. These were word reading fluency (unless the child was the one reading aloud), sex, rated attention span, introversion-extroversion, anxiety, and internal locus of control.

There were a total of five interactions (including a possible one) involving child characteristics. Four of these have already been mentioned as follows:

*The fluency of the child taking a turn as the oral reader interacted with teaching emphasis. This interaction was discussed at length in the section on homogeneous grouping. Figure 6 shows that the oral reader's fluency had especially strong negative effects when there was a stress on accurate word identification.

*The sex of the child interacted with the relative interest of the sentences to boys and girls. This interaction was reported in the section on materials. Figure 10 confirms that boys did better on boy-oriented sentences and girls did better on girl-oriented sentences.

*Reading comprehension ability interacted with the interest of the materials. This interaction was discussed in the materials section. Figure 9 shows that there was a greater difference between high and low ability children on high interest sentences than on low interest sentences.

*Finally there was a possible interaction between individual children's reading comprehension ability and the teaching emphasis. This interaction was discussed in the section on teaching emphasis. Figure 2 suggests that the higher the child's ability the greater the advantage of an emphasis on meaning.

One interaction not previously mentioned was also observed, involving a personality characteristic of the child, as follows:

*Introversion-extroversion interacted with position of the sentence in the lesson, $F(1,5445) = 8.93$, $p < .01$, % Var = .14, $B = .05$. Introverts did better early in the lesson while extroverts did better late in the lesson.

We have no explanation for this interaction. A general conclusion is that the yield from the noncognitive measures was low.

An Alternate Analysis of Sentence Recall

In order to be sure that the findings were invariant over scoring procedures, the reduced model was run with the lenient, gist recall score as the dependent variable. The mean percentage of recall that met lenient criteria was 38.6 and the standard deviation was 16.8. The findings were essentially the same as in the preceding analysis, though some effects became stronger, and some weaker. The following were the largest changes:

Reading comprehension ability had less effect, $F(1,244) = 21.89$, $p < .01$, % Var = 6.9, $B = 4.43$. The superiority of the meaning treatment

over the word identification treatment was greater, $F(1,244) = 20.37$, $p < .01$, % Var = 6.4, $B = 4.34$. Active turntaking made a larger difference, $F(1,8865) = 67.14$, $p < .01$, % Var = .69, $B = 4.39$. The fluency of the oral reader had a stronger negative effect, $F(1,8865) = 27.52$, $p < .01$, % Var = .28, $B = -1.25$.

The largest change between the two analyses was that Fry readability level now had a strong negative relationship with performance, $F(1,8865) = 62.74$, $p < .01$, % Var = .65, $B = 1.83$. The interpretation of this result is that when the requirement for a complete and precise answer is relaxed then the higher the grade level of the material the better is the recall.

Performance on the Word Identification Test

All of the results reported so far entail children's recall of sentences. Two other measures of lesson performance were the children's speed and accuracy in reading a list of 16 difficult words from the sentences. Number of errors and time to read the list in seconds were evaluated in analyses involving the treatment factors, individual difference factors (except the noncognitive factors), and all the two-way interactions of the treatment and individual difference factors.

The significant predictors of number of errors were the reading comprehension test, $F(1,244) = 212.28$, $p < .01$, % Var = 41.00, $B = -2.52$; the reading fluency test, $F(1,244) = 22.31$, $p < .01$, % Var = 4.31, $B = -.55$; group ability, $F(1,244) = 16.85$, $p < .01$, % Var = 3.25, $B = -2.49$; and the individual reading comprehension ability by group ability interaction, $F(1,244) = 22.49$, $p < .01$, % Var = 4.34, $B = .35$. The multiple correlation for the reduced model was .73. Whether a child received a

meaning emphasis or word identification emphasis had absolutely no effect on number of errors, $F < 1.0$, nor was the teaching emphasis involved in any interactions.

The significant predictors of time (in log seconds) to read the list were reading fluency, $F(1,244) = 1009.03$, $p < .01$, % Var = 78.20, $B = -.15$; group ability, $F(1,244) = 16.81$, $p < .01$, % Var = 1.30, $B = -.03$; and the reading fluency by type of teaching emphasis interaction, $F(1,244) = 7.11$, $p < .01$, % Var = .55, $B = -.01$. The multiple correlation for the reduced model was .90.

In instruction, like the rest of the world, you usually get what you pay for. The obvious prediction was that the word identification groups would do better than the meaning groups on the list of hard words. Whereas there was no difference among groups in number of errors, the obvious prediction was partially confirmed in the case of the time measure; as can be seen from Figure 11, children who were below average in reading fluency read the list of hard words faster if they had received a word identification emphasis. However, children of above average fluency read the list at about the same speed whichever treatment they had received. Thus, there was some evidence that receiving feedback on pronounciation and reading and rereading sentences improved automaticity for below average readers.

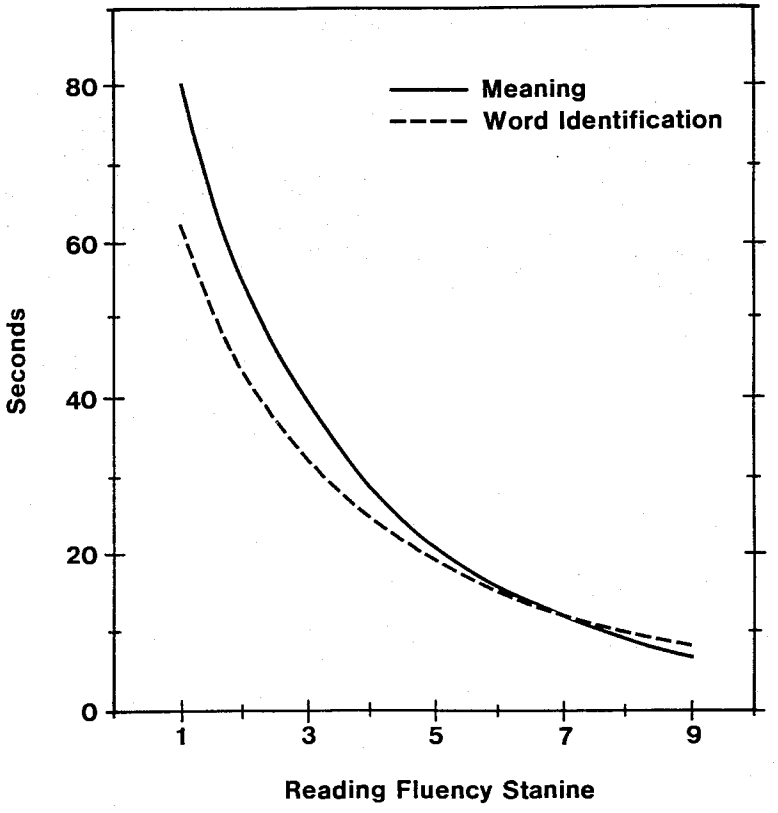
Insert Figure 11 about here.

Experiment 2

The main purpose of the second experiment was to see whether the principal findings of the first study could be replicated. In particular,

Figure 11

Time in seconds (backtransformed from log seconds) to read hard words as a function of teaching emphasis and reading fluency.



we were concerned to check the tenuous finding that the relative benefits of an emphasis on meaning depend upon the reading proficiency of the child. In the first experiment, an advantage for a meaning emphasis with less able children may have been obscured because of a performance floor. In the second experiment the list of sentences was shortened from 36 to 14 and the words were made easier. The average Fry readability level dropped from 4.3 to 2.0; its range was reduced from 7.0 to 4.5. As before, sentence interest was normed by another group of third grade children. No attempt was made to differentiate interest by sex this time.

Meaning was emphasized in the same way as in the first study. That is, for each sentence the child was asked to elaborate on the sentence--to provide a continuation that told what might happen next. In order to introduce a word identification emphasis, two difficult filler sentences, upon which most children made some miscues, were included early in the lesson. This gave the experimental teacher the opportunity to dramatize that he or she wanted accurate, fluent pronunciation of every word. Thereafter, almost all of the children were able to read almost all of the sentences with no trouble. Under both treatments, when a serious miscue or long hesitation did occur, the teacher supplied the word and the child kept going.

There were several other differences from the first study. Children participated individually rather than in groups. This was done to get an unconfounded assessment of whether individual reading ability interacts with teaching emphasis. Two control conditions were included in addition to the meaning emphasis and word identification emphasis treatments. These were a silent reading condition in which children were asked to read

carefully to themselves and a listening condition in which the sentences were played from an audio tape. Finally, the measure of performance, recall of sentences, was modified in order to leave no doubt that comprehension was being assessed. The children were tested for recall using cues related inferentially to the sentences contained in the lesson rather than words that had appeared in the sentences. For instance, for the sentence "The strong man chopped the wood," the cue was "axe." It is obvious that recall of this sentence would be unlikely unless it had been comprehended in the first place. The measures of speed and accuracy of reading difficult words from the sentences were omitted in this study because few hard words were used.

Design and Procedure

The subjects were 86 third graders from Homewood, Illinois. The mean national stanine for the sample was 6.34 and the standard deviation was 1.73. Thus, the reading proficiency of the sample ranged from poor to very good with more children above the national average than below. The children were taken from their classrooms one at a time to receive the experimental lesson. They were randomly assigned to the meaning emphasis condition, the word identification condition, the silent reading condition, or the listening condition.

After the lesson the children were asked to read the three lists of 15 words as quickly as possible, skipping those which they could not read. These were unrelated to the sentences used in the lesson, but yielded a measure of children's word reading fluency. Following this task the children were given a probe word, an unexpressed instrument or instantiation of each of the earlier 14 sentences (not including the difficult filler

sentences), and asked to tell the sentence it reminded them of. For instance, "The strong wind blew the roof off our house," the probe was the instantiation, "tornado." In the earlier example, the cue, "axe," is the unexpressed instrument.

The measure of reading comprehension was the comprehension subtest of the Metropolitan Achievement Test. It had been given by the classroom teachers five months before the study as a part of the school's regular testing program.

Results and Discussion

Overview of Analysis

To facilitate comparison with Experiment 1, a separate analysis of the meaning emphasis and word identification emphasis condition was done first. Then an analysis of all of the data was completed.

There were five between-subjects factors in the overall analysis. Entered first were reading comprehension ability and word reading fluency. Next came the treatment conditions entered as three orthogonal contrasts: TC1: Meaning emphasis versus the other three conditions; TC2: Listening versus word identification emphasis and silent reading; TC3: Word identification emphasis versus silent reading. The two-way interactions between the treatment conditions and reading comprehension ability and reading fluency were examined.

Within-subject factors included the Fry readability level and rated interest of the sentences. Also included were four characteristics of the words which served as cues for recall: whether the cue was related to the sentence by an instrumental inference or an inference of instantiation; a measure of the degree of association between the cue and the sentence; the

number of syllables in the cue; and frequency of usage of the cue. The latter three measures are of no intrinsic interest. They were included to prevent variations in the difficulty of the test items from confounding the effects of readability and interest. In this study the measure of the association of the cues and sentences was the number of children in the meaning emphasis group who included the cue word in their elaborations. All two-way interactions among the features of the cues and sentences and between these features and the treatments and facets of reading proficiency were examined. The logic for computing statistics and assessing significance was the same as in Experiment 1.

The interactions of treatment with other variables were computed by pooling the variance attributable to the interaction of each of the three treatment contrasts with the other variable for a single test of significance. This is a conservative procedure which acknowledges that the grouping of treatments was arbitrary (and that other possible groupings would have been as well). If any interaction computed in this manner had proved to be significant, this would have been regarded as a warrant to explore the relationship in detail, but none was significant.

In an alternate analysis, the Fry readability level of each sentence was replaced with the mean frequency of usage (the Carroll, Davies, & Richman, 1971, Standard Frequency Index) of the content words, the frequency of usage of the least frequent word, the total number of syllables in the sentence, and the mean number of syllables in content words. This analysis was somewhat more sensitive than the one using Fry readability level; however, none of the conclusions was different. Therefore, the analysis that

involved readability is reported here because the results will be understandable to a broader range of people.

Results

Meaning Emphasis Versus Word Identification Emphasis

The meaning treatment proved to have a substantial advantage over the word identification treatment, $F(1,37) = 55.63$, $p < .01$. The mean percentage of sentence recall was 71.1% in the former case but only 39.8% in the latter. It is especially noteworthy that in Experiment 2, unlike the first experiment, there was no hint of an interaction between the type of treatment and the standardized reading comprehension score of the child, $F(1,37) = 1.37$ nor between treatment and readability, $F < 1.0$. It should be cautioned, though, that this experiment did not provide as strong a test as Experiment 1 of possible interactions, because the range of individual differences in comprehension and, particularly, the range of readability were constrained. Nonetheless, it does seem reasonable to infer that there was a performance "floor" in Experiment 1.

An alternate interpretation of the difference between the two experiments is that the second one unconfounded individual and group ability. In the first experiment low ability individual children tended to be in low ability groups and low groups were allocated more time in which to learn the sentences.

The conclusion supported by both experiments is that a meaning emphasis gives better results with average and above average third-graders no matter

what the difficulty or interest of the materials, length of the task, or grouping arrangements. Experiment 2 suggests, in addition, that a meaning emphasis is also superior for poor readers, provided the task is within their range.

Comparison of All Treatments

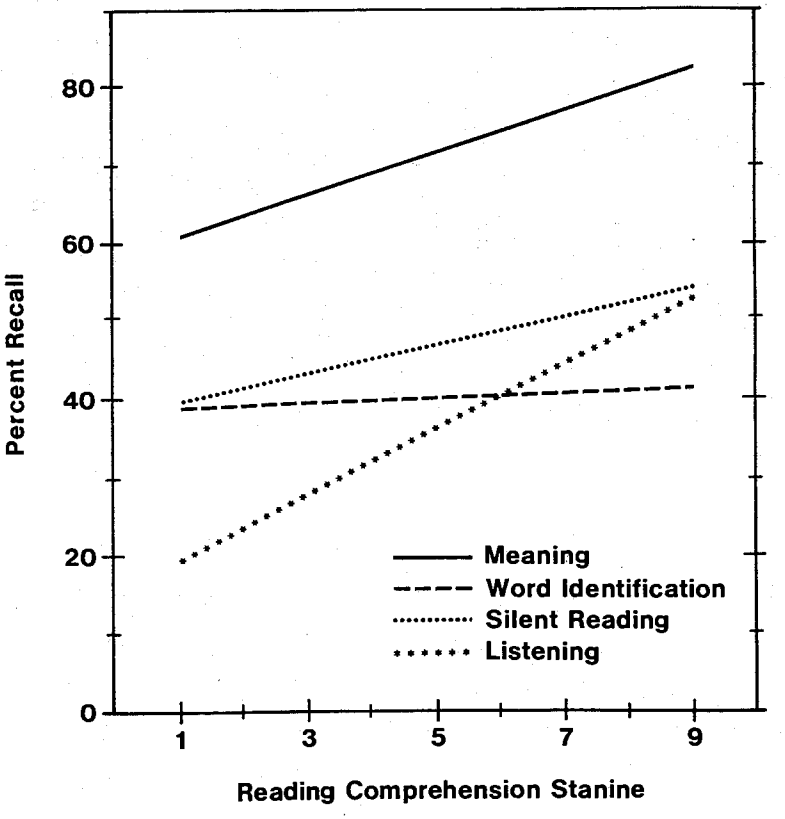
An average of 46.5% and 36.8% of the sentences were recalled in the silent reading and listening conditions, respectively. The figures for the meaning and word identification conditions, already reported above, were 71.1% and 39.8%. The meaning treatment gave much better results than the other three treatments, $F(1,75) = 58.13$, $p < .01$, % Var = 38.6, $B = 22.6$, whereas the other three treatments did not differ, $F(2,75) = 2.19$, $p > .10$.

Figure 12 portrays sentence recall under each of the four treatments as a function of the reading comprehension stanines of the children. This figure is included in order to allow an easy comparison with the results of Experiment 1 graphed in Figure 2. Figure 12 may appear to show an aptitude-treatment interaction, but it was not significant, $F(3,75) = 1.63$, $p > .10$. The most striking fact revealed by Figure 12 is that the very poorest readers who received the meaning emphasis treatment performed better than the very best readers who received any of the other three treatments.

Insert Figure 12 about here.

Figure 12

Percentage recalled under each of the four treatments as a function of reading comprehension.



Materials

The aspects of materials that predicted recall were, first of all, the association between the cue word and the rest of the sentence, $F(1,1093) = 77.86$, $p < .01$, % Var = 6.01, $B = 1.03$. The second significant predictor was another aspect of the cue, the frequency of usage of the cue word, $F(1,1093) = 47.03$, $p < .01$, % Var = 3.63, $B = -1.37$. The more rare the word the better cue it made, a result consistent with the findings that Schnorr and Atkinson (1970), among others, have obtained with adult subjects.

Interest had a strong relationship with performance, $F(1,1093) = 57.34$, $p < .01$, % Var = 4.43, $B = 69.84$. The quadratic of interest was also significant, $F(1,1093) = 14.96$, % Var = 1.16, $B = -23.25$. Interest had approximately the same relationship with performance as it did in Experiment 1.

Fry readability level did not influence performance, $F(1,1093) = 2.79$, $p > .05$, % Var = .22, $B = 2.20$. In fact, the direction of the relationship was wrong; there was a slight tendency for supposedly harder sentences to be recalled better. Interest and the quadratic of interest accounted for 26 times as much variance as readability. The poor showing of readability in this study is perhaps not surprising, since all of the sentences were rather easy and the range was constricted. Nonetheless, the findings of the two experiments together leave no doubt that interest is much more important than readability.

There was no significant interactions in Experiment 2.

Summary and General Discussion

The major question this research was designed to answer was whether a meaning emphasis or a word identification emphasis gives better results in a third grade reading lesson. The answer was unequivocal. A meaning emphasis produced better results. Even in the first experiment, in which there was a complicating interaction with the readability of the materials and a possible interaction with the reading proficiency of the children, there were very few sentences and very few children with which a word identification stress worked better. Furthermore, the second experiment suggested that these interactions were probably spurious. The only clear advantage for the word identification emphasis that appeared anywhere in this research was that poor readers who got this emphasis were able to pronounce hard words from the lesson faster than poor readers who received the meaning emphasis.

From the evidence on differential treatment of children in different ability groups, which Hiebert (1982) has summarized in an excellent recent review, it appears that low groups typically get an instructional regimen that resembles the word identification emphasis used in the present research. Whereas no children are getting much comprehension instruction (Durkin, 1978-79; Mason & Osborn, 1982; Nielson & Rennie, 1981), Hiebert has summarized evidence showing that there is more attention to meaning in high groups. The present research does not indicate any need for nor advantage from differential treatment of high and low groups. Though regrettably Experiment 1 left some loose ends, both the preferred analysis of the data from this experiment and the data from Experiment 2 support the

conclusion that a meaning emphasis gets better results than a word identification emphasis with poor readers as well as good readers.

Perhaps one reason there is so little comprehension instruction in the typical classroom, particularly for poor readers, is that the schools operate on the presumption that once children have learned to decode they will comprehend by doing what comes naturally. This is a dangerous presumption, one which is inconsistent with the data presented here, particularly the data from the second experiment. If "doing what comes naturally" were enough, the high ability children, at least, would have performed much better in the silent reading and listening conditions. The fact that instructions to elaborate on the presented sentences had a large effect on the performance of children of every ability level shows that the children were not spontaneously engaging in meaningful processing.

Other recent research points to a similar conclusion. Notably Hansen and Pearson (1982) found that second graders and fourth graders recalled dramatically more from basal reader-type stories when preparation for reading included questions and discussion designed to activate schemas appropriate for understanding the story content. The advantage was especially large for poor readers. From this, Hansen and Pearson inferred that good readers have already developed partially satisfactory comprehension strategies and spontaneously use them some of the time, whereas most poor readers either don't have the strategies or don't use them unless prompted. It would appear from the Hansen and Pearson research that poor readers are in special need of instruction that will facilitate comprehension. Yet these are just the children whose attention is invariably directed to surface features of language.

While we do not know for sure, it is a reasonable conjecture that constant attention to the surface of language will interfere with the development of a persistent tendency to read deeply. Getting meaning from any text beyond the simplest basal story requires understanding an imposing array of linguistic devices and the development and consistent use of a variety of strategies for monitoring comprehension and organizing and integrating information (see Collins & Smith, 1982). It stands to reason that poor readers are going to have a hard time developing these competencies if their teacher always focuses their attention on the sounds words make.

Educators are likely to be concerned, not only with whether effects are significant, but with whether they are large enough to be of practical value. The average advantage of a meaning emphasis in the first experiment was not large, but this had more to do with how the experiment was calibrated than with the intrinsic power of the treatment. The second experiment showed that when the task is within the range of the children, a meaning emphasis has a huge advantage over a word identification emphasis. In fact the poorest readers who received a meaning emphasis did better than the most able readers who received any other treatment. Thus, this research confirms that the Law of Meaningful Processing holds for children under conditions that simulate classroom reading instruction.

The practical educator is also likely to worry that the teaching emphases employed in these experiments were more extreme than teachers would actually use, and that a good teacher would work on both word identification, when needed, and meaning. Our reason for studying what might be regarded as extremes is that, when previous data in an area are

inconclusive and opinions are in conflict, it is often useful to design an experiment as one would write an opera: exaggerate characteristics beyond those that may be seen in nature; examine the limiting case; and take care to observe the full range of every variable. In a murky area, premature study of blends and shades may add to confusion about fundamental issues. The value of knowing for sure what will happen under limiting conditions ought not to be underestimated.

The practical educator may also be concerned because the conclusions reached in this research are based on recall of material in one day's lesson. What would happen in the long run? Could it be that there are latent benefits to a word identification emphasis that require a big investment in drill and practice in order to become manifest? We don't know the answer, but until someone finds out for sure a good rule of thumb is this: in general there is no reason to believe that an ineffective treatment that loses in the short run will become a winner in the long run if it is repeated day after day, week after week.

Finally, we must address a question that will be on the minds of concerned members of the lay public as well as professional educators. What does this research have to say about the teaching of reading at the earliest stages? The answer is "very little." In particular, it has little bearing on whether systematic, direct instruction in phonics should be a component of a beginning reading program. Of course, there is every reason to suppose that the Law of Meaningful Processing holds for six- and seven-year-olds as well as eight- and nine-year-olds and adults. Furthermore, it seems highly probable that predominant attention to letter-sound correspondences would compete with meaningful processing, if only because

instructional time is limited. This insight is not original nor is it especially controversial; it is a point that would be acknowledged by any reasonable advocate of systematic phonics instruction. The controversy is about whether, once letter-sound principles are under control, children will turn their main attention to meaning, as is hoped by the advocate of direct instruction, or whether, instead, once subjected to direct phonics instruction, many children become trapped in a futile echo chamber of disembodied sounds, as is feared by the advocate of the position that letter-sound patterns should be acquired as the incidental byproduct of efforts after meaning. The research reported here does not illuminate this issue.

We turn now to a consideration of the other findings of the research. These will be treated more briefly. We will comment on just the highlights.

The interesting finding about group ability was that it was negatively related to sentence recall in the word identification groups. A two-part theory was proposed to explain this finding: (a) the lower the ability of the group, the more instructional time the teacher allocated and (b) increased instructional time led to greater learning. With the data in hand, it was possible to prove that this simple theory provides a quite satisfactory explanation. However, another theory probably gives a better, more refined explanation. The augmented theory places the reading fluency of the child taking the active turn, instead of group ability, on the causal path leading to enhanced sentence recall. Indirect evidence showed the plausibility of the augmented model; however, available data did not permit a direct, quantitative test of its adequacy.

Probably neither of these models would give a good account of the dynamics of learning in the typical classroom. Evidence from classroom

research suggests that teachers usually fix the amount of time that will be allowed each group rather than fix the amount of material that will be covered, as the experimental teacher in these studies did. Putting this another way, our experimental teacher allowed more time to compensate for the slower pace in low groups. Apparently most teachers do not do this. Therefore, neither model will explain children's performance in the typical classroom. More important, though, than whether any particular model provides a good account of classroom dynamics is the demonstration here and in other recent research (Leinhardt, Zigmond, & Cooley, 1981) that it is possible to formulate precise models of aspects of reading instruction and subject these models to rigorous, quantitative evaluation.

That children's mastery of the day's lesson did not depend on amount of instructional time when meaning was stressed is a most newsworthy finding. It is consistent with the findings of basic research. At one time, when experimental psychologists mainly studied lists of words and nonsense syllables, it was thought that learning inevitably depended upon time (Cooper & Pantle, 1967). Subsequent research suggests that this belief is false, most especially when the learner is induced to engage in deep semantic processing (Craik & Watkins, 1973).

To conclude that learning from written material is not a time-dependent process when the learner is engaged in an effort after meaning is not to imply that no time is required. Every reading process takes some time and the time it takes depends, at least, upon the skill of the reader and the length of the material. The fact that instructional time was unrelated to sentence recall when there was a meaning emphasis suggests that even the poorest readers had enough time for processing in the fastest

moving of the meaning-oriented groups. Therefore, the poor performance of the poor readers must have been due to ineffective processing; it could not have been due to time limitations because, again, if it were, there would have been a positive relationship between instructional time and performance under the meaning emphasis condition. On the other hand, if poor readers generally had enough time, then presumably good readers usually had more time than the minimum they needed. In traditional theories, learning is characterized as a process of accretion. The strength or probability of learning is assumed to increase continuously as a function of the time invested in rehearsal or elaboration. But no theory of this sort will explain our results because the "strength" of the sentence representations of the good readers in slow moving, meaning emphasis groups would be predicted to increase, and if this were happening instructional time would have been positively correlated with sentence recall. We propose the alternative theory that once the time requirements of preliminary processes, such as perceptual analysis, are satisfied the learning of the basic propositional units expressed in sentences is an all-at-once event triggered by the integration of meaning (see Auble, Franks, & Soraci, 1979; and Goetz, Anderson, & Schallert, 1981). After this event has happened, further time is redundant.

Probably most practicing educators will not be surprised by the fact that active turntaking has a fairly large effect. Perhaps what is surprising is that, to the best of our knowledge, no previous researcher has succeeded in demonstrating the benefits of an active role. In keeping with the philosophy of the experiment-as-opera, a strictly-ordered turn-taking procedure was employed in this research so that it would be crystal

clear to the children when it was and was not their turn. It stands to reason that more open turntaking procedures would cause more children to realize the benefits of active involvement more of the time. In support of this hypothesis, Au and Mason (1981) have recently reported that native Hawaiian-Polynesian children show higher rates of engagement during a lesson when the children can enter the discussion whenever they wish, and overlapping and joint turns are permitted, than when the teacher enforces one-at-a-time turntaking.

Earlier we pointed out that the failure to find an interaction between the teaching emphasis and interest must mean that these two factors have their influence on different underlying facets of the reading process, for if the influence were on the same facet then the effects of the two factors together would be redundant. Exactly the same logic applies to the failure to find interactions between active participation and the teaching emphasis and active participation and interest. Thus this research makes the prima facie case that these important factors affect different stages or aspects of the reading process.

The interestingness of the materials was a very powerful variable in these experiments. It was more important by an order of magnitude than readability, the criterion everywhere used to gauge the appropriateness of school materials. Indeed, the relationship between readability and performance was not even in the right direction in either experiment. In the worst case, involving the lenient gist scoring of sentence recall in Experiment 1, each grade level increase in "difficulty" on the Fry readability scale was associated with a 1.83% increase in sentence recall. Sentences rated at the seventh grade level, which supposedly are too

difficult for third graders, were recalled 12.8% more often than sentences rated at the first grade level. It is important to understand why this happened and what it means. Artifactual explanations probably can be ruled out. Sentences with a wide range of readability were used. Several other important aspects of the materials were manipulated in a design that made them approximately orthogonal to readability. The method of analysis insured that readability got just the weight it deserved other things being equal.

Readability formulas have been sharply criticized in recent years (e.g., Davison, 1982). A chief complaint is that none of the formulas takes account of the interrelationships among sentences and, therefore, text that lacks cohesion or clear organization gets passed off as "readable" by children in a certain grade. Proponents of the readability formula tend to answer their critics obliquely, dismissing objections on the grounds that the facts indicate that the formulas "work."

We wish to briefly sketch the claim that the facts that allegedly support the use of readability formulas are suspect because they are based on research that contains errors of design, analysis, and interpretation. A problem of analysis was mentioned earlier. Readability research has always followed the dubious practice of aggregating data across children, computing the performance of the mean second grader, the mean third grader, and so on for each of a wide range of text selections. Then the mean performance of the children at the different grades is predicted from features of the words and sentences in the texts using regression analysis. The basis for the claim that readability formulas "work" is that analyses of this type often explain 80% to 90% of the variance. However, few people

are curious to explain the performance of the hypothetical mean third grader. What most educators and behavioral scientists really want is to understand diverse individual third graders. If this is the goal, then the proper analysis involves the whole array of answers of every child on every text, as we explained earlier. If the right analysis were done, the amount of variance explained would fall precipitously and the reading field would have a proper amount of humility about how well readability formulas actually work.

A more damaging criticism of readability formulas is that they provide only superficial indicators of the deeper reasons for reading ease or difficulty. The major factor is whether the reader possesses background knowledge adequate for assimilating a text. We hypothesize that vocabulary difficulty, the principal component of every formula, is primarily a proxy measure for background knowledge. Anderson and Freebody (1981; see also Freebody & Anderson, in press) used an example from the jargon of sailing to convey this point. A child who knows the meaning of the word "spinnaker" is quite likely to be able to understand sentences that do not even contain "spinnaker," such as, "The sloop jibed suddenly and the boom snapped across the cockpit." Obviously the underlying factor that facilitates or inhibits comprehension of this sentence is extent of knowledge of sailing. The general point is that the jargon that goes with a topic is just the tip of the conceptual iceberg.

In the studies in which readability formulas were validated, high-flown language always has been associated with subtle, abstract knowledge, simple language with everyday themes. Over the years, the association probably has grown stronger because heavy controls on the readability of

children's texts have made the language in them unnaturally simple. This means that the confounding of knowledge demands and language complexity has been exacerbated in the recent studies (e.g., Bormuth, 1966; Coleman, 1970) in which people place most faith. The result is that the formulas now in use egregiously overestimate the importance of surface features of language. Probably most third graders could get the gist of a story about a girl and her puppy even if it were dressed up in fancy language, whereas no amount of simplification of the language of an economics treatise would permit very many third graders to grasp the concept of the multiplier effect. If the foregoing arguments are correct, it is not surprising that readability had weak effects, trending negative, in the two studies reported in this paper, since the theme of every sentence was easily grasped by third graders.

Finally, in the research in which readability formulas were validated children read without the help of a teacher. Ironically, readability controls are heaviest, not in library books that children do read by themselves, but in basal readers. These are almost always read under the direct supervision of a teacher who can help with difficult words. Under the word identification treatment in the present research, a mistake on a difficult word, far from being a problem, was an opportunity for learning (see Figure 1).

In addition to whatever strengths it may have, every program of research, indeed, every approach to research, has its limitations. We grow increasingly weary of the quarrelsomeness of proponents of educational research styles who spend more time attacking other styles than they do providing answers to interesting questions using the style that they are

touting. In mature areas of inquiry, everyone accepts that there must be converging evidence from a variety of sources before important propositions can be regarded as proved. In these areas, it is well-understood that chains of inference are long and that it is seldom that any one discipline has the tools to forge every link. In this spirit, we readily acknowledge the limitations of the research reported here. The principal limitation is that the research lacks ecological validity: an unfamiliar teacher taught especially constituted groups of children; sentences were used instead of texts; the measures involved mastery of the material in one day's lesson. This was basic instructional research which ought not to be generalized directly to regular classroom instruction. We have research in progress attempting to test the major conclusions from the studies reported here in natural classroom settings.

On the other hand research of the kind reported in this paper should not be undervalued. In these experiments, it was possible to study treatments rarely seen in nature, for instance, lessons that were both meaning oriented and intensive and systematic. It was possible to study the independent effects of variables that are correlated in nature, for instance, the ability of the group and the relative emphasis the teacher gives to meaning and word identification. It was possible to get a preliminary look at the causal dynamics in reading groups, a difficult if not impossible task in a naturalistic investigation. Probably the most distinctive feature of the present research was that it was designed to investigate the interplay of a large number of the factors that converge at given moments to determine whether particular sentences will be mastered or not. Because of this, it was possible to demonstrate that the child

taking a turn was getting more than the other participants; and, it was possible to make more subtle discoveries, such as that the benefits of active turntaking and the interestingness of sentences do not extend to the immediately following material. Principally, though, it was possible to show that the Law of Meaningful Processing continues to operate over a diverse range of conditions and with both good and poor readers.

References

Allington, R. L. Teacher interruption behaviors during primary grade oral reading. Journal of Educational Psychology, 1980, 72, 371-374.

Anderson, R. C. Allocation of attention during reading. In A. Flammer & W. Kintsch (Eds.), Discourse processing. Amsterdam: North Holland Publishing, 1982.

Anderson, R. C., & Freebody, P. Vocabulary knowledge. In J. Guthrie (Ed.), Comprehension and teaching. Newark, Del.: International Reading Association, 1981.

Anderson, R. C., Goldberg, S. R., & Hidde, J. L. Meaningful processing of sentences. Journal of Educational Psychology, 1971, 62, 395-399.

Anderson, R. C., & Hidde, J. L. Imagery and sentence learning. Journal of Educational Psychology, 1971, 62, 526-530.

Anderson, R. C., & Kulhavy, R. W. Learning concepts from definitions. American Education Research Journal, 1972, 9, 385-390.

Asher, S. R. Topic interest and children's reading comprehension. In R. J. Spiro, B. C. Bruce, & W. F. Brewer (Eds), Theoretical issues in reading comprehension. Hillsdale, N.J.: Erlbaum, 1980.

Au, K. H. & Mason, J. M. Social organizational factors in learning to read. Reading Research Quarterly, 1981, 17, 115-152.

Auble, P. M., Franks, J. J., & Soraci, S. A. Effort toward comprehension: Elaboration or "aha!"? Memory and Cognition, 1979, 7, 426-434.

Baddeley, A. D. Short-term memory for word sequences as a function of acoustic, semantic, and formal similarity. Quarterly Journal of Experimental Psychology, 1966, 18, 362-365.

Baddeley, A. D. The trouble with levels: A reexamination of Craik and Lockhart's framework for memory research. Psychological Review, 1978, 85, 139-152.

Beckerman, T., & Good, T. L. The classroom ratio of high- and low-aptitude students and its effect on achievement. American Education Research Journal, 1981, 81, 317-327.

Berliner, D. C., & Rosenshine, B. The acquisition of knowledge in the classroom. In R. C. Anderson, R. J. Spiro, & W. E. Montague (Eds.), Schooling and the acquisition of knowledge. Hillsdale, N.J.: Erlbaum, 1977.

Bormuth, J. R. Readability: A new approach. Reading Research Quarterly, 1966, 1, 79-132.

Carroll, J. B., Davies, P., & Richman, B. Word frequency book. Boston, Mass.: Houghton Mifflin, 1971.

Chall, J. S. Learning to read: The great debate. New York: McGraw-Hill, 1967.

Collins, A., & Smith, E. E. Teaching the process of reading comprehension. In D. K. Detterman & R. J. Sternberg (Eds.), How and how much can intelligence be increased. Norwood, N.J.: Ablex, 1982.

Cooper, E. H., & Pantle, A. J. The total time hypothesis in verbal learning. Psychological Bulletin, 1967, 68, 221-234.

Craik, F. I. M., & Lockhart, R. S. Levels of processing: A framework for memory research. Journal of Verbal Learning and Verbal Behavior, 1972, 11, 671-684.

Craik, F. I. M., & Tulving, E. Depth of processing and the retention of words in episodic memory. Journal of Experimental Psychology: General 1975, 104, 268-294.

Craik, F. I. M., & Watkins, M. J. The role of rehearsal in short-term memory. Journal of Verbal Learning and Verbal Behavior, 1973, 12, 599-607.

Crandall, V. Intellectual achievement responsibility questionnaire. Princeton, N.J.: ETS test collection, Microfilm No. 006098, 1975.

Davison, A. Readability--Appraising text difficulty. In R. C. Anderson, J. Osborn, & R. Tierney (Eds.), Learning to read in American schools. Hillsdale, N.J.: Erlbaum Associates, Inc., 1982.

Duffy, G. G. Teacher effectiveness research: Implications for the reading profession. In M. Kamil (Ed.), Directions in reading research and instruction. Washington, D.C.: National Reading Conference, 1981.

Durkin, D. What classroom observations reveal about reading comprehension instruction. Reading Research Quarterly, 1978-79, 14, 481-533.

Emerson, R. W. Essays. New York: Literary Classics, 1945.

Eysenck, H., & Eysenck, S. Eysenck junior personality questionnaire. San Diego: Educational and Industrial Testing Service, 1975.

Freebody, P., & Anderson, R. C. Effects of vocabulary difficulty, text cohesion, and schema availability on reading comprehension. Reading Research Quarterly, in press.

Fry, E. Fry's readability graph: Clarification, validity, and extension to level 17. Journal of Reading, 1977, 20, 242-252.

Glover, J. A., Brunning, R. H., & Blake, B. S. Distinctiveness of encoding and recall of text material. Journal of Educational Psychology, 1982, 74, 522-534.

Goetz, E. T., Anderson, R. C., & Schallert, D. L. The representation of sentences in memory. Journal of Verbal Learning and Verbal Behavior, 1981, 20, 369-385.

Goodman, K. S., & Goodman, Y. M. Learning to read is natural. In L. B. Resnick & P. A. Weaver (Eds.), Theory and practice of early reading (Vol. 1). Hillsdale, N.J.: Erlbaum, 1979.

Hansen, J., & Pearson, P. D. An instructional study: Improving the inferential comprehension of good and poor fourth-grade readers (Tech. Rep. No. 235). Urbana: University of Illinois, Center for the Study of Reading, March 1982.

Hiebert, E. H. An examination of ability grouping for reading instruction. Reading Research Quarterly, 1982, in press.

Hill, K. Motivation, evaluation and educational testing policy. In L. J. Fyaus, Jr. (Ed.), Achievement motivation. New York: Plenum, 1980.

Hyde, T. S., & Jenkins, J. J. The differential effects of incidental tasks on the organization of recall of a list of highly associated words. Journal of Experimental Psychology, 1969, 82, 472-481.

Jacoby, L. L., Craik, F. I. M., & Begg, I. Effects of elaboration of processing as encoding and retrieval: Trace distinctiveness and recovery of initial context. Journal of Verbal Learning and Verbal Behavior, 1979, 18, 585-600.

Jenkins, J. R., Pany, D., & Schreck, J. Vocabulary and reading comprehension: Instructional effects (Tech. Rep. No. 100). Urbana: University of Illinois, Center for the Study of Reading, August 1978.

Just, M. A., & Carpenter, P. A. A theory of reading: From eye fixations to comprehension. Psychological Review, 1980, 87, 329-354.

Kane, J. H., & Anderson, R. C. Depth of processing and interference effects in the learning and remembering of sentences. Journal of Educational Psychology, 1978, 70, 626-635.

Leinhardt, G., Zigmond, N., & Cooley, W. Reading instruction and its effects. American Education Research Journal, 1981, 18, 343-361.

Mason, J. Overgeneralization in learning to read. Journal of Reading Behavior, 1976, 8, 173-182.

Mason, J. M. A description of reading instruction: The tail is wagging the dog (Reading Ed. Rep. No. 35). Urbana: University of Illinois, Center for the Study of Reading, August 1982.

Mason, J. M., & Osborn, J. When do children begin "reading to learn?": A survey of classroom reading instruction practices in grades two through five (Tech. Rep. No. 261). Urbana: University of Illinois, Center for the Study of Reading, September 1982.

Nielson, A. R., & Rennie, B. Conditions affecting reading to learn in content area classrooms. Paper presented at the National Reading Conference, Dallas, 1981.

Paris, S. G., & Lindauer, B. K. The role of inference in children's comprehension and memory for sentences. Cognitive Psychology, 1976, 8, 217-227.

Pearson, P. D., Gallagher, M., Goudvis, A., & Johnston, P. What kinds of expository materials occur in elementary school children's textbooks. Paper presented at the National Reading Conference, Dallas, 1981.

Perfetti, C. A., & Lesgold, A. M. Discourse comprehension and sources of individual differences. In M. Just & P. Carpenter (Eds.), Cognitive processes in comprehension. Hillsdale, N.J.: Erlbaum, 1977.

Perfetti, C. A., & Lesgold, A. M. Coding and comprehension in skilled reading and implications for reading instruction. In L. B. Resnick & P. A. Weaver (Eds.), Theory and practice of early reading (Vol. 1). N.J.: Erlbaum, 1979.

Pflaum, S. W., Walberg, H. J., Kanegianes, M. L., & Rasher, S. P. Reading instruction: A quantitative analysis. Educational Researcher, 1980, 9(7), 12-18.

Ross, B. H. The more, the better? Number of decisions as a determinate of memorability. Memory and Cognition, 1981, 9, 23-33.

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.

Schnorr, J., & Atkinson, R. C. Study position and item differences in the short- and long-term retention of paired associates learned by imagery. Journal of Verbal Learning and Verbal Behavior, 1970, 9, 614-622.

Stein, B. S., & Bransford, J. D. Constraints on effective elaborations: Effects of precision and subject generation. Journal of Verbal Learning and Verbal Behavior, 1979, 18, 769-777.

Tharp, R. G. The effective instruction of comprehension: Results and description of the Kamehameha Early Education Program. Reading Research Quarterly, 1982, 17, 503-527.

Venezky, R. L., & Massaro, D. W. The role of orthographic regularity in word recognition. In L. B. Resnick & P. A. Weaver (Eds.), Theory and practice of early reading (Vol. 1). Hillsdale, N.J.: Erlbaum, 1979.

Watts, G. H., & Anderson, R. C. Effects of three types of inserted questions on learning from prose. Journal of Educational Psychology, 1971, 62, 387-394.

Footnote

[1]The research reported herein was supported by the National Institute of Education under Contract No. US-NIE-C-400-76-0116. The authors gratefully acknowledge the cooperation of the third grade teachers and children from Urbana, Ill., Homewood, Ill., and Rantoul, Ill. who participated in this research. We also are pleased to acknowledge the important contributions made by Taffy Raphael, Paul Wilson, William May, William Tirre, William Nagy, and Alexis Freeman. Some of the material contained in this paper was presented in an address at the International Reading Association convention in Chicago, May 1982.

[2]The authors are indebted to Ali Iran-Nejad for suggesting this rival hypothesis.

This page is intentionally blank.