ILLINOIS

UNIVERSITY OF ILLINOIS AT URBANA-CHAMPAIGN

PRODUCTION NOTE

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

Technical Report No. 490

HOW DISCUSSION QUESTIONS INFLUENCE CHILDREN'S STORY UNDERSTANDING

Linda G. Fielding
University of Iowa at Iowa City

Richard C. Anderson
P. David Pearson
University of Illinois at Urbana-Champaign

January 1990

Center for the Study of Reading

TECHNICAL REPORTS

UNIVERSITY OF ILLINOIS AT URBANA-CHAMPAIGN
174 Children's Research Center
51 Gerty Drive
Champaign, Illinois 61820

CENTER FOR THE STUDY OF READING

Technical Report No. 490

HOW DISCUSSION QUESTIONS INFLUENCE CHILDREN'S STORY UNDERSTANDING

Linda G. Fielding
University of Iowa at Iowa City

Richard C. Anderson
P. David Pearson
University of Illinois at Urbana-Champaign

January 1990

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

The work upon which this publication was based was supported in part by the Office of Educational Research and Improvement under Cooperative Agreement No. G0087-C1001-90 with the Reading Research and Education Center. The publication does not necessarily reflect the views of the agency supporting the research.

Abstract

This research examined the influence of discussion questions with a dual focus on important and implied text ideas in children's story understanding. In a three-week instructional study, questions with such a focus formed the basis of the treatment for the four experimental groups, while control group questions came from the basal series teachers' manuals containing the experimental stories. Questions for each experimental group differed in focus (prediction or review) and in whether they required answer justification. Comprehension of 106 third-grade children randomly assigned to the five groups was measured with free and probed story recalls. Results of mixed hierarchical analyses indicated that justifying answers significantly improved the performance of groups asked prediction questions but not groups asked review questions. There was also a nonsignificant trend toward superior performance of the experimental groups over the control group, suggesting that importance and inexplicitness of ideas are appropriate foci in designing discussion questions.

HOW DISCUSSION QUESTIONS INFLUENCE CHILDREN'S STORY UNDERSTANDING

When Frank Smith (1975, p. 34) called comprehension "the condition of having one's cognitive questions answered," he was using *questions* as a generic term for a whole set of activities and cognitions that skilled readers seem to engage in--predicting, reviewing, making inferences, and modifying interpretations, to name a few. Many researchers and practitioners, in fact, seem to view teachers' questions as a means of modeling and engaging children in various active processes that a skilled reader carries out independently to make meaning while reading. The key to making questions useful in that way may be asking systematically derived and strategically placed questions whose purpose and pattern eventually might be internalized by children--because, after all, the ultimate goal of comprehension instruction is student independence in reading, not giving answers to teacher-posed questions. The present study is an attempt to integrate two lines of research about comprehension instruction to devise a method of generating and asking discussion questions about complex stories. The purpose is to study the role of discussion questions with a dual focus on important and implied text ideas in children's comprehension of discussed stories and stories that they read independently.

Related Literature

The present study fits into a body of research about the nature of text and teacher questions and their effects on comprehension that has grown considerably, especially in the last decade. Classroom observations of comprehension instruction (Durkin, 1978-79; Guszak, 1967; Mason & Osborn, 1982; Morrison, 1986) and analyses of the instructional suggestions in basal reader teachers' manuals (Beck, McKeown, McCaslin, & Burkes, 1979; Durkin, 1981) have documented the preponderance of story questions; the emphasis they place on low-level, factual information while failing to develop plot or event sequences through lines of questions about stories; and the tendency to ask questions that may distract readers from a story's central content.

Such descriptions of the state of classroom comprehension instruction have spurred a number of investigations into the changes in instruction that might lead to more desirable learning outcomes. Consistent with current views of reading as a schema-theoretic process (e.g., Adams & Collins, 1979; Anderson & Pearson, 1984; Rumelhart, 1977, 1980), these investigations highlighted the role of background knowledge in comprehension--both in its role of providing a prototypical organizing framework, a story schema for example, that enables readers to anticipate and organize text ideas (e.g., Fitzgerald, 1984; Fitzgerald/Whaley, 1981; Stein & Glenn, 1979), and in its contribution of the raw material out of which crucial story inferences can be made (e.g., Paris & Lindaeur, 1976; Pearson, Hansen, & Gordon, 1979). With respect to the role of questions, most notable are two general findings: (a) both training and practice in answering inference questions can improve not only children's ability to answer inference questions but also their general understanding of what they read (Gordon & Pearson, 1983; Hansen, 1981; Hansen & Pearson, 1983; Raphael & McKinney, 1983; Raphael & Pearson, 1985; Raphael & Wonnacott, 1985; Redfield & Rousseau, 1981; Sundbye, 1987); and (b) attention (through questions or otherwise) to the typical structure of stories (e.g., Fitzgerald & Spiegel, 1983; Gordon & Pearson, 1983; Idol, 1987; Singer & Donlan, 1982) or to the information otherwise identified as central or important in stories (Beck, Omanson, & McKeown, 1982; Omanson, Beck, Voss, & McKeown, 1984) can improve comprehension.

The present study addresses several gaps in our knowledge about the role of story discussion questions in comprehension instruction: (a) the lack of integration between our knowledge about the structure of stories and our knowledge about inferential reading comprehension; (b) our incomplete knowledge about the effectiveness of a special class of inference questions, those that require predictions; and (c) our failure to test systematically the common-sense notion that it is beneficial for children to justify their answers to discussion questions with evidence from the text or their background knowledge.

These issues are important in planning effective text-based instruction that centers on discussion questions.

First, although the value of both central or important story questions and inferential questions has been documented, the relationship between the two has not been investigated. Even in stories for young children, what is stated in the surface structure usually gives only a partial account of the complex relationships one must recognize to understand and appreciate the story. For example, characters' goals, the relationship between their goals and their actions, and the conflict between two or more characters' goals often are left to inference, although they provide the locus of causality in stories (Bruce, 1980a, 1980b, 1984; Bruce & Newman, 1978). Yet it is not clear how any of the most prevalent story structure theories represent inferences or account for their role in the comprehension of complex narratives. Furthermore, a close look at the questions in the redesigned story lessons of Beck and her associates (Beck et al., 1982; Omanson et al., 1984) shows that while all questions tapped central story content, less than one-third required inferring even the most straightforward connections between different sentences, and only half of those required integration of information in noncontiguous sentences or integration of text information with background knowledge. The general approach to potentially difficult text ideas in the work of Beck et al. (1982) seems to be to supply the difficult ideas (e.g., that raccoons, because of their masks, might look a little like bandits) instead of leading children to figure them out through questions and discussion.

On the other hand, it is not clear how theories of inferential reading comprehension explain which inferences, out of the almost limitless number that could be generated about a complex story, are the most important ones for story comprehension. Trabasso and his colleagues (Nicholas & Trabasso, 1980; Omanson, 1982; Trabasso & Nicholas, 1980; Trabasso, Secco, & Van Den Broek, 1984; Warren, Nicholas, & Trabasso, 1979) suggest that out of all the inferences that could be made while reading a story, the ones most critical are those necessary for the comprehension of an event chain in a narrative. The degree to which inference instructional studies have focused on such inferences is unclear. In Hansen's (1981; Hansen & Pearson, 1983) studies, for example, important text ideas were the basis for inferences in the group taught an inference-making strategy, but only a strategy for what Warren et al. call slot-filling inferences was taught. The training strategy for making inferences was based on selecting important text ideas, then having students predict what would happen in the story from their prior experience in similar situations. For Hansen's inference question-only group (the group that practiced answering a lot of inference questions), it is not clear whether any inferences required what Warren et al. (1979) call text connecting nor whether the inference questions tapped important text ideas. Gordon and Pearson (1983) and Raphael and her colleagues (Raphael & McKinney, 1983; Raphael & Pearson, 1985; Raphael & Wonnacott, 1985) taught strategies for making both text-connecting and slot-filling inferences, but it is unclear what role importance played in the selection of examples for instruction or in the construction of inference questions.

Besides this key issue of no research with a combined focus on importance and explicitness of the story ideas that are targeted with questions, the present research also addresses the issue that no studies have investigated the relative effects on comprehension of review and discussion questions. Although predicting is a characteristic of good readers (e.g., Collins, Brown, & Larkin, 1980; Fitzgerald, 1984; Fitzgerald/Whaley, 1981), a special case of inference-making (Pearson & Johnson, 1978), and an especially active and independent form of engagement with the text (e.g., Stauffer, 1975), few investigators have involved children in prediction through questions. The successful strategies of Hansen (1981; Hansen & Pearson, 1983) and Palincsar and Brown (1984) incorporated prediction. Furthermore, recent work by Anderson and his associates (Anderson, Wilkinson, Mason, & Shirey, 1987; Wilkinson, Anderson, & Wilson, 1987) showed that prediction questions produced better memory for stories than did word-level questions when each was asked exclusively throughout small-group reading lessons. However, neither these investigators nor any others have carried out the more stringent test of comparing prediction and review questions that tap the same important story ideas.

Finally, no research attention has been devoted to the value of having children justify their answers to questions with evidence from the text and from background knowledge. Although this practice frequently is advocated in reading methods textbooks (e.g., Mason & Au, 1986; Stauffer, 1975), it is supported at present only by conventional wisdom. Giving evidence for answers seems to be a useful way to engage children in reviewing and integrating text information and background knowledge, noticing and revising any incorrect or implausible answers, and making thought processes public, but no studies have compared the performance of children required to justify their answers with children who are not required to justify their answers.

Overview of the Present Study and Hypothesized Outcomes

A logical step in this line of research, then, was to select stories that require making text-connecting and slot-filling inferences for comprehension of the central story information, and to examine the effects on comprehension of several manipulations of discussion questions. In the present study, questions for each experimental group placed special emphasis on making the important inferences necessary for understanding the story's central event sequence as depicted by a map of the story (derived from a combination of methods of representing story information). Story comprehension of children in these experimental groups was compared with the comprehension of children who were asked story discussion questions suggested in basal reader teachers' manuals. Finer differences in the instruction of the four experimental groups reflected three related questions of instructional interest: whether prediction and review questions have differential effects on story comprehension, whether consistently justifying answers has a more positive effect on comprehension than not justifying answers, and whether the relative effect of justifying answers is mediated by whether questions require prediction or review of story ideas. Table 1 summarizes this description of the differences among groups in the study.

[Insert Table 1 about here.]

Story comprehension was measured before, during, and immediately after a three-week instructional study. Lesson comprehension measures included oral free and probed recall of the stories read and discussed on the 4th, 8th and 12th days of instruction. Pretest and transfer measures included oral free and probed recall of stories read independently without discussion questions. Free recalls were scored for recall of important and unimportant explicitly stated story propositions and for plausible inferences. Probed recalls included literal, inferential, and thematic/story extension questions.

The following group differences were hypothesized:

1. The four groups who were asked story map-derived questions were expected to perform better than the group asked basal reader questions on free recall of important literal information, generation of plausible inferences in free recall, and all probed recall questions--literal, inferential, and thematic/story extension.

The reasoning behind this hypothesis was that discussion questions derived from a story map, because they came from an examination of the story's structure and degree of explicitness, would highlight important story information, follow the story's sequence, and encourage deeper processing (e.g., Craik & Lockhart, 1972; Jacoby, Craik, & Begg, 1979; Glover, Bruning, & Plake, 1982; Glover, Plake, Roberts, Zimmer, & Palmere, 1981; Glover, Plake, & Zimmer, 1982) through special emphasis on making important inferences. Although thematic/story extension questions were not a special focus in story discussions, it was reasoned that deeper processing would encourage richer elaborations that would lead to a greater likelihood of producing plausible answers to these questions.

2. The two groups required to justify their answers during discussions were expected to perform better than the two groups not required to justify their

answers on the same measures--important literal recall, plausible inferences in recall, and all probe questions.

The reasoning behind this hypothesis was that answer justification would serve to increase interaction with both literal and inferential story information, facilitate revising of incorrect or implausible answers, and encourage students to examine the thought processes that led them to their answers. Again, although thematic and story extension questions were not the focus of instruction, it was reasoned that deeper processing and examination of thought processes would enhance the elaborations that would produce more plausible answers to these questions.

3. No hypothesis was made about the overall effectiveness of prediction versus review questions, because it was believed that under certain conditions prediction alone, without justifying or verifying answers, could be ineffective. If students spontaneously reason through their predictions, incorporate text ideas into their hypotheses, and reconsider their predictions after further reading, they may outperform students who are asked review questions. However, if available story information is not used to form reasoned predictions, and if predictions that do not match story outcomes are not revised after further reading, students might remember what they predicted rather than what actually happened in the story, with detrimental effects on all recall measures. Therefore, whether there is a net effect for prediction over review was expected to depend largely on whether students monitor their predictions spontaneously in the predict-only group.

Method

Subjects

Data were collected in two phases to obtain a statistically adequate sample size in each of the five treatment groups. A total of 106 third-grade children participated in the study, 58 in Phase 1 (February-March 1986) and 48 in Phase 2 (April-May 1986). The three schools involved (one with three third-grade classes in Phase 1, two with one third-grade class each in Phase 2) were from two middle-sized Illinois cities. The original plan was to block students on their standardized reading comprehension test scores and then randomly assign one student from each block to each of the five treatment groups, but practical constraints necessitated some deviation from this plan.

First, the stipulation was added that no children were assigned to a treatment group that met during the time their regular classroom reading group normally met with the teacher or any time slot during which they normally attended a special pull-out class. Second, since Phase 2 was conducted in two separate schools, two treatment groups were run in one school and three in the other. Finally, a schedule change forced one child in the Phase I predict-plus-justify group to switch to the review-plus-justify group after only one day of instruction; and after one week of instruction in Phase I, two children in the review-only group dropped out of the study so they would not miss a new computer class.

The total number of students in each of the instructional groups across both phases of the study thus ranged from 19 to 24. The grand total of 106 children represents all but eight available children across the two phases of the study: two from Phase I who did not have parental permission to participate, two from Phase I who dropped out, and three from Phase I and one from Phase 2 whose achievement test scores and teachers' judgment suggested that they would be frustrated by the instructional materials. Table 2 shows the number of children and the average total reading percentile per group in each school.

[Insert Table 2 about here.]

Instructional Materials

The same nine stories, chosen from third- and third/fourth- grade level basal readers not used in the participating schools, served as the basis for the 12 lessons in each group. Differences among group treatments were in the way the stories were divided into silent reading units--segments of the text that children read and discussed before further silent reading--and in the discussion questions asked before or after each silent reading unit.

Stories. All of the stories were structured around a problem-solving schema--in other words, each was based around a problem that a character tried to solve or interacting problems that several characters tried to solve. In addition, all required making text-connecting or slot-filling inferences for understanding some categories of story information, and all were either 500-800 or 1,000-1,600 words in length so they could be read and discussed in either one or two 30-minute sessions. Stories included realistic and fanciful tales, both humorous and serious, both folktales and other genres.

Story maps. Generation of discussion questions that tapped important story content depended on creating a map of each story that represented the important literal and inferential story information. The investigator devised a procedure for representing the relatively long and complex stories used in this study that made use of existing story grammar and causal chain representations that had been applied to simpler stories (e.g., Stein & Glenn, 1979; Trabasso et al., 1984). Briefly, explicitly stated story information was divided into story categories typically used in story grammar analysis--*setting(s)* plus episodes, each of which might include an *initiating event*, which set up a problem one of the characters had to solve; the *goal* of this character; one or more *plans* or *attempts* the character makes to solve the problem; and the *result(s)* of and *reaction(s)* to each attempt. Because the stories were long and often complex, not every story proposition was mapped--for example, neither details that could be subsumed under a more inclusive term nor information that did not advance the main story line represented by the problem-solution framework was mapped.

Second, categories of information that were not stated explicitly in the text, or that were disclosed out of sequential order in the story, were filled in and marked as inferences. Implied causal connections between categories of information similarly were filled in and marked as inferences. Potential sources of comprehension difficulty thus could be located and targeted with questions later.

For example, a story called "The Birthday" (Hart, 1977) concerns a girl named Delilah who is almost 10 years old (setting) who knows exactly what she wants for her birthday, but can't tell her parents because if she does it will not be a surprise, and her parents love surprises (initiating event of Episode 1). The story continues,

> At dinner that night, she dropped a hint.
> "We're studying how sound works," she said. "We learned how you get sound out of a record player. It's very, very interesting."
> "How?" her father asked.
> Delilah told him. (Plan 1 of Episode 1)
> You're right," he said. "That's very, very interesting." (Result 1 of Episode 1) (Hart, 1977, p. 62).

Several mapping conventions can be illustrated with this excerpt from the story. First, in the first paragraph of the story, besides the information that Delilah is almost 10 years old, the story lists six extraneous facts about Delilah (e.g., "She's the tallest girl in her class. She plays basketball and drums.") that are *not* mapped because they do not advance the main story line. Second, the story does not explicitly state here or elsewhere that what Delilah wants for her birthday is a record player. The reader must infer this from hints Delilah drops in various plan sequences throughout the story. This inference is an important one if the reader is to understand Delilah's surprise at the end of the story

when she receives a puppy instead of a record player for her birthday. Thus, in the story map representation, it would be listed as part of Delilah's goal and would be marked as an inference. Finally, it is only in this first plan sequence that Delilah's hint is explicitly signalled (with the words "At dinner that night, she dropped a hint."). In the rest of the story, the reader must infer her motivation in doing things like stopping in a department store to look through the record bins when she goes shopping with her parents, asking her father what his favorite kind of music is in the car on the way home, and informing her mother that she has borrowed a record from the library and is going over to her friend's house to listen to it. A further complexity of this story is that the parents' thoughts are left entirely to inference. The reader is left to decide, later in Episode 2, whether Delilah's parents consciously elect to get her a gift other than the one she hinted for, or whether they simply fail to "get" the hints she gives in Episode 1.

Silent reading units. The resulting map for each story first was used to divide the story into silent reading units--places where it would make sense to stop reading and monitor comprehension, which was accomplished in this study with discussion questions. The decision to ask questions during story reading rather than mass them at the end of the story reflects the present study's goal of modeling for young readers what it is that skilled readers do to achieve understanding. Skilled readers probably do not wait until they have finished reading a whole story to decide whether they have understood it; rather, they probably monitor their understanding in the course of reading. The logic behind exactly where the silent reading unit boundaries were placed was based partly on the story map and partly on whether prediction or review questions were to be asked.

Silent reading units for the two review question groups ended where some sort of resolution had been reached about part of the story. These breaks sometimes occurred at the ends of episodes, sometimes where an embedded episode began, and sometimes in multiple-plan episodes after the character tried one plan and before he or she tried another. Logical breaks for the two groups being asked prediction questions were places in the story where students had enough information to make reasoned guesses about what was going to happen next, but before closure was reached. These places might be during the initiating event, before the problem was stated; after the goal, before the plan was stated; after the plan, before the result was clear; after the result, before the character's reaction was clear; between plan sequences in an episode; between episodes; or at the end of the story.

For example, in the story "The Birthday," the first silent reading unit for prediction groups ended after the initiating event was described--thus, children predicted how Delilah might solve her problem before reading about any of her plans. Review groups, however, read all three of the plans that Delilah believed had been unsuccessful before engaging in any discussion. For the basal reader control group, the silent reading units suggested in the teacher's manual were used.

Each child in each group read from a specially prepared copy of the story that contained only one of their group's silent reading units per page, with the word *STOP* written at the end of the silent reading unit. Students in each group typically read 250-300 words in each silent reading unit, after which the investigator asked the discussion questions appropriate for the group.

Discussion questions. Discussion questions for each experimental group were based on two interrelated considerations--centrality and inexplicitness of text ideas. Emphasis was placed on asking questions that tapped central story information that also had to be inferred. Questions for the review groups recapitulated the important information in a given silent reading unit and integrated it with previously read information. Questions could be asked about the literal information needed to make important inferences; about missing categories of information, particularly if the character's goal was unstated; about implied connections between different categories of information; and about implied connections between episodes.

Questions for the prediction groups required predicting upcoming important information that the reader could make reasoned guesses about at that point in the story. Questions were designed so that

for the most part, the information the review groups used for their answers was the same information that the prediction groups were making hypotheses about.

Members of the review-plus-justify group also explained how they arrived at their answers using background knowledge and available text information. Members of the predict-plus-justify group, in addition to justifying their answers in this way, also compared their predictions to actual text events after further reading.

A control set of questions from each story came from the basal reader teacher's manual.

The same excerpt from "The Birthday" that has been discussed already will be used to illustrate the types of discussion questions asked in each group. Review groups were asked the following questions about the part of the story described above:

1. What does Delilah want for her birthday?
2. Why couldn't she tell her parents exactly what she wanted for her birthday?
3. How did she first hint that she wanted a record player?

Prediction groups, after reading through what is identified above as the *initiating event* of Episode 1 were asked these questions:

1. What do you think Delilah might want for her birthday?
2. What do you think Delilah could do to let her parents know what she wants for her birthday?

The basal reader teacher's manual suggested asking these four questions about the part of the story described above:

1. What birthday is Delilah celebrating? How do you know? Read the sentence that gives you a clue.
2. How would you describe the way Delilah looks?
3. How does Delilah try to get her message across to her parents without telling them?
4. Do you think her father understood her hint? Why or why not?

Notice that the review groups make one important inference and paraphrase two pieces of central explicitly stated information in answer to their questions. The prediction groups' questions engage them in predicting, before they read it, the information that the review groups paraphrase after reading. The group asked basal reader questions, in addition to one important inferential question (number 4 above), are asked one question whose answer is a literal detail (number 1), one whose answer actually may distract them from central story information by focusing on extraneous information outside of the information depicted in the story map (number 2), and one rather cryptic question that skirts around but evades the important implied idea that what Delilah wants is a record player (number 3).

The number of questions per session for the experimental groups was held to a maximum of 9 or 10 to allow adequate time for discussion. Because responses to the basal reader questions generally took less time, 15 or 16 questions per session were planned. Basal manual questions were not reviewed by the experimenter until after experimental questions (and test questions for lesson comprehension measures) were prepared.

Measures

Story tests. All test materials were prepared by the experimenter. Children recalled and answered probe questions about three instructional stories spaced throughout the study (called *lesson comprehension measures* in this study) and about pretest and transfer stories selected and mapped according to the same criteria used to select and map the instructional stories.

For the pretest and transfer stories, oral free and probed recalls were collected after the students read the story silently with an examiner present. Students were allowed to ask for word pronunciations while reading, but they read independently and did not discuss these stories. For three instructional stories--those read on the 4th, 8th and 12th day of instruction--students orally retold and answered probe questions about the story on the day after instruction.

For all five oral free recalls (the three lesson comprehension measures, the pretest, and the transfer task), students retold the story to one of four or five adult examiners. To encourage students to give complete free recalls, including inferences, they were asked to tell the story into a tape recorder so that a first or second grader who had not read the story would be able to understand even the hard parts of the story after listening to their tape. Further attempts to elicit inferences included prompting the students once to tell more after they had finished their recall, and, in a manner similar to that used by Tierney, Bridge, and Cerra (1978-79), prompting with up to two general questions that asked students to tell more about literal information they had reported but not made inferences about.

Probe recalls for each of the tested stories (three instructional stories, the pretest and the transfer story) consisted of 15 questions. Five questions tapped important literal story information, five required important text-connecting or slot-filling inferences, and five (called thematic/story extension questions in this study) went beyond the text's story line in some way by asking about the story's theme or message, children's evaluations of certain characters, or alternative story outcomes. Students had to recombine text information and make judgments related to it to answer these questions. Although this study was concerned directly with the more text-based aspects of comprehension, it was important to know whether variations in text-based discussions would affect more divergent responses such as the ones tapped with these thematic/story extension questions.

For example, for the story "The Birthday," a literal probe question was, "What was Delilah's problem in letting her parents know what she wanted for her birthday?" An inferential probe was, "What was the real reason that Delilah told her parents about how sound comes out of a record player at dinner one night?" A thematic/story extension question was, "Why do you think Delilah's parents gave her a puppy instead of a record player for her birthday?"

Affective response measure. A possible corollary effect of the experimental treatments was assessed with an affective response measure. Students' affective responses to the stories, which it was believed might be more strong and positive in the prediction conditions, were measured using a story affect questionnaire with the format of and most of the same questions as a questionnaire used by Jose and Brewer (1984) with children. Students responded to 10 questions such as "How much did you like this story?" or "How interesting did you think this story was?" by checking one of seven boxes arranged from very small ("a little") to very large ("a lot"). This measure, which took only a few minutes to complete, was collected for each of the five test stories.

Procedures

Lessons. The experimenter, an experienced classroom reading teacher, conducted all group lessons. Every day, the procedure was the same. All groups had the same short vocabulary/background knowledge introduction of only a very few words or concepts whose understanding was central to the story.

Students in the basal reader control group then read the story or each silent reading unit (depending on the teachers' manual directions) for the purpose stated in the teachers' manual. For instance, for the story "The Birthday," the purpose for reading the whole story was, "The title gives us a good idea of what the story may be about. Read to find out what the girl wants for her birthday and what happens on that day. Pay special attention to the way she lets her parents know what gift she wants and think about what you would do if you were this girl."

Each experimental group, in contrast, had a global strategy-related reading purpose that was the same each day. The strategy-related purpose for each group matched that group's experimental manipulation. On the first day of instruction, the experimenter briefly described how stories do not always tell the reader everything, so the reader has to think during the whole time that she or he is reading. Review groups were told to think about how all the pieces of the story fit together with each other and about why the characters do the things they do. In addition to that, the review-plus-justify groups were told to think about what information they used to fill in the missing pieces or to figure out how the pieces of the story fit together--they might use things they already knew in their lives, things they already knew about stories, and things they had already read in the story. The strategy-related purpose for prediction groups was to use what they already knew and what they had already read to keep trying to figure out what would happen next in the story. Predict-plus-justify groups also were told to think about what information they used to make their predictions, and to compare their predictions with what actually happened in the story after more reading.

These discussions took less than 5 minutes on the first day of instruction. After that, at the start of each lesson the experimenter asked children to recall what they had been doing every day in their reading lessons. The experimenter repeated salient points from the children's responses and added points they omitted.

Silent reading and discussion in basal reader groups followed a format typical for small-group lessons. After students had read a section of the story silently, several questions were asked. For each one, the experimenter asked the question, children bid for a turn to answer by raising hands, and the experimenter selected a respondent. When the respondent's answer was correct or plausible, the experimenter said "Right" and unless another student volunteered to say something else, the experimenter went on to the next question. Multiple responses were not sought actively. When an answer was incorrect or implausible, the question was asked again and again until a correct response was given. Respondents were not asked to explain their answers unless the teacher's manual specifically directed the teacher to do so. For example, a basal reader question about "The Birthday" was, "How did Delilah feel about her new puppy?" Any plausible answer was accepted without a request for explanation because the teacher's manual did not direct the teacher to ask anything like, "How could you tell?"

In the experimental groups, in contrast, multiple responses were sought actively. For every question accompanying a silent reading unit, the experimenter asked the question, gave children time to mentally frame an answer, then called on one respondent (a record was kept so all students were asked for an initial response an approximately equal number of times). In the review groups, without judging the first answer, the experimenter then asked if anyone wanted to say anything more or anything different. The experimenter tried to stay out of the role of answer evaluator by giving group members a chance to decide whether answers were implausible or incomplete and by helping students to see that two different answers could sometimes be "right." Occasionally the experimenter had to push for a plausible answer when no one in the group disagreed with an implausible one.

In the prediction groups, all students were encouraged to commit themselves to a prediction. In the predict-plus-justify groups, students also explained how they formed their predictions, and, after further reading, compared their predictions to the author's choice for how to continue the story.

Scoring free recalls. To score children's free recalls, a propositional representation of each story was produced and compared to the story's map to identify each important literal, unimportant literal, and important implied proposition. Then each child's free recall was matched to this propositional representation. Procedures are described in detail below.

The story map representation of each story was used as the basis for free recall scoring decisions. First, each story was divided into clauses, each of which counted as one story proposition. Additionally, each new piece of information about the time, place, or main characters of the story (i.e., the *setting* information) counted as a proposition, even if it did not comprise a separate clause in the story. Thus, in "The Birthday," for example, the clause "Delilah Bush is almost 10 years old" counted as two propositions, not one, because that clause introduced a main character for the first time ("Delilah Bush") and established a setting detail that is related to the central event sequence of the story ("is almost 10 years old").

Once all stated story propositions were so listed, important slot-filling and text-connecting inferences--those that filled in missing categories of story information or missing links between categories of information--were added to the propositional representation. For example, the idea "Delilah wants a record player" was added to the list of propositions, even though it was implied and not stated in the story, because it explicates what the story map identified as her *goal*.

Finally, stated story propositions were classified as "important" or "unimportant" using the following scheme: Any proposition that contained new story category information was counted as important, while other information was unimportant.

The propositional representation of the above excerpt from "The Birthday" will be used to clarify how these decisions were made.

016 At dinner that night she dropped a hint.
017 "We're studying how sound works," she said.
018 "We learned how you get sound out of a record player.
019 It's very, very interesting."
020 (Because she wants them to know she is interested in record players.)
021 "How?" her father asked.
022 Delilah told him.
023 "You're right," he said.
024 "That's very, very interesting."
025 (He doesn't seem to get the hint.)

Important and explicitness decisions were as follows:

016 and 018 are literal and important, because each explains part of Delilah's first plan.

017 and 019 are literal and unimportant, because they embellish the plan, but add no new information.

020 is implied and important, because it explicates the connection between Delilah's plan and her goal.

021 and 022 are literal and unimportant, because they add no new information.

023 and 024 are literal. Either *one* is important, because each partially explains the result of this plan. Because the important idea in these propositions is redundant, only one per free recall is counted as important. If both are present, the other counts as unimportant.

025 is implied and important, because it further explains the result of this plan and suggests why Delilah will have to come up with another plan.

This method of representing story information was decided upon because it represents important implied ideas as well as literal ones, and it differentiates important from unimportant information while still providing a mechanism for giving children some credit for everything they remember.

To match a child's recall to the propositional representation of the story, the child's recall was divided into units that represented as many separate propositions of the story as possible, and any elaborations on the story. For example, one child's recall of the story segment represented as propositions 016-025 above was the following:

1 And what she wanted for her birthday was a record player.
2 And the first hint she gave them was at the dinner table.
3 She said that she'd been studying sound in school
4 and that she . . . thought the sound was pretty
5 and that she'd like a radio maybe--I mean a record player.

This part of the recall was scored in the following way:

1 was matched to an earlier proposition, 011, and counted as *important implied.*
2 was matched to 016 and counted as *important literal.*
3 was matched to 017 and counted as *unimportant literal.*
4 was counted as *unimportant implied.* It is an elaboration that neither advances nor interferes with story comprehension.
5 was counted as *incorrect information*--Delilah did not tell her parents she wanted a record player; she only hinted it.

In addition to *important literal, unimportant literal, important implied, unimportant implied,* and *incorrect* information, a child's recall could include several other types of information. A description and an example of each from the story "The Birthday" is given below.

Subsuming Statement: any inclusive statement that subsumes two or more propositions.

Example: She gave her parents hints about four times.

Explanation: This child inferred a common feature of the four plan sequences in the story--each is a hint to Delilah's parents. For the major analyses, these were added to *important inferences.*

Evaluative/Judgmental Comment: The child gives his opinion about the quality of the story or about a character or event.

Examples: It was a good story.
The ending was dumb.
She should have told them she wanted a record player.

Evaluative/judgmental statements were so infrequent that they were not considered further in any analyses.

Thus, four categories of free recall were of special interest: *important literal, unimportant literal, important inference,* and *elaborative inference.* Incorrect information was kept track of for separate consideration but was not examined in any major analyses. Each child received a score in each category that summed the number of unprompted propositions he recalled with the new propositions he added after the prompts he was given during his oral free recall.

The experimenter scored all free recall protocols. A randomly selected 10% of the protocols from each story were rescored by a graduate student majoring in reading, after the scoring system was explained to her. Interrater reliability for each of the stories on the four major recall categories (important literal, unimportant literal, important inference, elaborative inference) ranged from .84 to .89.

Scoring probed recalls. Answers to literal questions were scored as correct (1) or incorrect (0). Answers to inferential and thematic/story extension questions were scored as plausible (1) or implausible (0). Partial credit could be given if a question asked for answer justification and the child did not supply it. For example, the question "What kind of girl is Delilah? How can you tell?" was awarded 1/2 point if the child offered a plausible description of Delilah's character or personality but did not, upon prompting, tell why she thought so. Occasionally, children's answers represented alternative story interpretations. Individual decisions were made about the plausibility of these interpretations. Proportion correct out of 5 for literal, inferential and thematic/story extension questions and total proportion correct out of 15 questions were computed.

The examiner scored all probed recalls, and one of three graduate students rescored each one. Initial interrater reliabilities ranged from .87 to .95. All scoring discrepancies were resolved through discussion between the experimenter and the other scorer.

Handling of missing data. Although complete data was available on the standardized reading test and was obtained through makeup testing when necessary for lesson comprehension and transfer measures, children sometimes were absent for instruction on a story and therefore could not be tested on it. These children, in spite of generally good attendance, would have been eliminated from the repeated measures analyses (which require complete data on the repeated measure) if their missing data were not filled in. Eliminating 20 children from this study simply because they missed one day of instruction was not acceptable; therefore, steps were taken to fill in missing data with the best possible estimate of what the child's scores would have been if he or she had been present for instruction and taken the lesson comprehension tests.

Briefly, separate regression equations predicting performance on each instructional story were calculated for the students in each group for whom complete data were available. These equations used all available data--the school, the standardized reading test total reading percentile, the total mean score on the story pretest, and the total mean score on the other two instructional stories besides the one whose score was being predicted. Once these equations were computed, they were used in predicting individual students' missing values.

Group means for each question type and recall type, as well as total mean scores, were compared for only the children with complete data and for the total group including any filled-in data. Because differences were negligible, only the data in which estimated values have been substituted for missing data are reported throughout this study.

Results

Each of the four main analyses described below--lesson comprehension free recall, lesson comprehension probed recall, transfer free recall, and transfer probed recall--used a mixed design with one between-subjects factor, one or two within-subjects factors, and three covariates. All analyses were performed using the MANOVA program in the *SPSS-PC+* package (Norusis, 1986a, 1986b).

The between-subjects portion of each analysis was carried out in the same way, so it will be described here only once. First the covariates, then treatment group contrasts, and finally covariate-by-treatment interactions were entered. School district was entered first in an attempt to control for differences between districts that were not of interest in this study--mainly, the different standardized reading tests used. Total reading percentile from the district's standardized reading test was entered next, followed

by the child's total score from the appropriate story pretest measure (oral free recall or oral probed recall, depending on the dependent measure in the analysis). Because these pretest measures are of general interest throughout the analyses, group means on them are reported separately in Table 3.

[Insert Table 3 about here.]

Four orthogonal contrasts representing hypothesized treatment group differences were entered into the between-subjects analysis after the three covariates. Contrast 1 (the story map contrast) compared the basal reader group with all four experimental groups to determine the relative effects of basal reader and story map-derived discussion questions. Contrast 2 (the justification contrast) measured the effects of answer justification by comparing the two no-justification groups with the two justification-of-answer groups. Contrast 3 (the question focus contrast) measured the relative effects of prediction and review questions by comparing the two prediction groups with the two review groups. Contrast 4 assessed the justification-by-question focus interaction to determine if the relative effect of justifying answers was mediated by whether prediction or review questions were asked.

The three covariate-by-treatment interactions were entered into the between-subjects part of the analysis last--school by treatment, total reading by treatment, and story pretest by treatment. None of these interactions was significant in any analysis, so interpretation will focus on the treatment group contrasts controlled for the covariates.

The within-subjects part of each analysis varied slightly, so each one will be described in detail when it is reported. When appropriate, tests of simple main effects or post hoc comparisons followed the overall analyses.

Tests of homogenity of error variance were carried out routinely in the *SPSS-PC+* MANOVA program. Although in the lesson comprehension measure analyses some subscores violated the univariate homogeneity of error variance assumptions, the original mixed analyses were carried out nonetheless for a more complete picture of the results. The between-subjects part of each analysis, upon which the major interpretations are based, uses what amounts to group grand means, which did not violate the assumptions. Nor were homogeneity of error variance assumptions violated in the analyses of transfer effects.

Because the treatment group contrasts are orthogonal, the significance of each can be examined separately even if the treatment effect is not significant overall. Since group differences are the main focus of interest in this study, the separate treatment group contrasts instead of the overall treatment effect are reported in the between-subjects ANOVA summaries. However, because there were no significant separate treatment contrast-by-covariate interactions and just one significant treatment contrast-by-within subject factor interaction that is described elsewhere, only the overall interactions involving treatment are reported in the ANOVA summary tables for economy of presentation.

Lesson Comprehension Measures Analyses

Free recalls. The ANOVA summary table for the overall analysis of free recall lesson comprehension measures is shown in Table 4. Because there were no interactions between the treatment contrasts and the within-subjects part of the design, group treatment effects are interpreted directly. The treatment group means presented in Table 5 represent the mean total number of recalled propositions (important literal + unimportant literal + important inferential + elaborative inferential) collapsed across the three stories.

[Insert Tables 4 & 5 about here.]

Of the four orthogonal contrasts for treatment, only Contrast 4, the justification-by-lesson focus interaction, is significant ($F(1,86) = 6.54, p = .01$), indicating that the effect of justifying answers was

different in the prediction and review groups. Post hoc comparisons of the two prediction groups and the two review groups indicated that while answer justification led to significantly better performance in the predict-plus-justify group than the predict-only group (difference in means = 5.7, $p<.05$), the means of the two review groups were much closer (difference in means = 1.3) and slightly, though nonsignificantly, favored the no-justification group. Possibly there was more of a press for plausibility of predictions and resulting attention to text ideas in the predict-plus-justify group than in the predict-only group, whereas the general focus on inference-making in the review groups led to deeper processing of text ideas in both review groups, even without a special focus on answer justification. These results are interpreted further in the discussion.

Although the main focus of this study is on treatment effects, results from the within-subjects part of the design are presented for a more detailed picture of ANOVA findings. The within-subjects analysis included two repeated measures factors--one for story (Story 1, Story 2, Story 3) and one for recall type (important literal, unimportant literal, important inferential, elaborative inferential). The reader is referred again to Table 4 for a summary of the within-subjects ANOVA results.

The significant recall type-by-story interaction ($F(6,606) = 172.47, p < .001$) indicates that the relative difference in recall type mean scores depends on the story. However, because the same basic trend is evident in each story, the separate cell means are not reported here. The significant main effect for recall type ($F(3,303) = 714.82, p < .001$) and the pattern of recall type scores summed across the three stories shows that important literal information is recalled more than any other type (important literal--16.5, unimportant literal--2.2, important inferential--3.0, elaborative inferential--2.9). Since there was more important literal information to recall than any other type, this is a completely expected result.

Probed recalls. Table 6 shows the ANOVA summary table for the overall analysis of probed recall lesson comprehension measures, and Table 7 shows observed means, means adjusted for the covariates, and standard deviations for each question type within each story. The within-subjects part of the design included two repeated measures factors--one for question type (literal, inferential, thematic/story extension) and one for story (Story 1, Story 2, Story 3). The significant treatment-by-question type-by-story interaction ($F(16,404) = 1.82, p = .026$) indicates that group effects depend on certain combinations of question type and story. Interpretation, then, will focus on this interaction. Although it could have been examined in several ways, it was the group differences that were of primary interest, so the a priori treatment group contrasts were tested for significance for each separate combination of story and question type. So trends of these results can be observed more easily, the results of these contrasts are presented in Table 8.

[Insert Tables 6, 7, & 8 about here.]

The first notable result is that for literal probes in the first two stories but not the third, the four experimental groups (whose story discussions centered on story map-derived questions) performed on average better than the basal group (Contrast 1). In Story 3, the one story where this effect was not observed, an examination of means shows high performance for all groups on the literal probe questions (range of adjusted means = .80-.89). There is some trend, then, toward better performance from the experimental groups than the basal group on literal questions. Explanation of this expected result is straightforward. Any time the experimental groups were asked literal questions during story discussions, the questions highlighted important information; moreover, all probed recall questions that were literal highlighted important information. In contrast, although the proportion of literal discussion questions in the basal reader discussion groups was high, the focus of the questions was not consistently on important information.

The second notable result, consistent with the overall results of the free recall analysis, is again an indication that justification of answers is differentially important in review and prediction groups (Contrast 4). Here, the effect is evident in performance on inferential probe questions, where the justification-by-question focus interaction was significant in two stories and close to significant (p =

.08) in the third. A closer examination of this interaction indicated that while answer justification led to significantly better performance in the prediction treatments (difference in adjusted mean proportion correct for inference probes across all three stories = .193; $F(1,86) = 18.42, p = .000$), the means of the two review groups on inferential probes were not significantly different (within .04 of each other on every story). Apparently the emphasis on inferential questions in the review groups was sufficient to engage the readers in deeper processing of the text ideas without the added emphasis on answer justification and verification. After all, to give plausible answers to inferential questions that are in some way text dependent, one must interact with the text information. It is possible, in contrast, that the nature of the discussions in the prediction-only group actually fostered departure from the text ideas and hence a tendency to discount these ideas in inference-making. This possibility will be explored further in the discussion.

The third notable trend in these results is that all significant and near-significant effects are confined to literal and inferential questions. There is no indication of group differences on thematic/story extension questions. Although it had been hypothesized that a focus on important ideas and on making inferences also might result in better performance on these more divergent questions (e.g., "What do you think the point of this story is?" "What kind of person would you say _____ is?", etc.), this does not appear to be the case.

One exception occurred, though, with the subset of these probes that asked for a new story ending or an alternative action for the protagonist. Because this sort of question was asked regularly at the end of story discussions in both prediction treatments, it was of interest to know whether there were group differences on this subset of probe questions. Five of the thematic/story extension questions across the three lesson comprehension measures fit this description--e.g., "Pick one of the animals Anansi the spider captured. Tell another way you think he could have captured it, and why." or "What do you think Delilah will do a couple of weeks before her next birthday? Why?" or "Do you think Stuey will ever do his homework in invisible ink again? Why/Why not?" When group means were examined on just these five responses across the three instructed stories, it was found that the same general trend that was apparent in other parts of the data appeared here as well: The predict-plus-justify group performed better than all other groups, and the predict-only and basal groups were lowest in the rank ordering (predict plus justify--.78, review plus justify--.71, review--.70, basal--.64, predict--.64).

Trends. Although these free and probed recall lesson comprehension measure analyses were directed by the a priori treatment contrasts, examination of treatment group means reveals other consistent trends that support the original hypotheses with some modifications. Most important, in some cases it was the unexpectedly low performance of the predict-only group that kept the experimental groups' performance from surpassing that of the basal reader control group. When the predict-only group is not considered, the other three experimental groups' average performance is significantly better than the basal control group's performance on literal and inferential probes across all stories (literal: basal--.72, others--.82, $F(1,86) = 12.4, p < .001$; inferential: basal--.65, others--.74, $F(1,86) = 6.45, p = .013$).

Furthermore, when the focus is put on the most stable measures--general results collapsed across all stories and types of free or probed recall--a general pattern emerges in the rank ordering of groups, with the predict-plus-justify group outperforming all others, the two review groups performing nearly the same and next in the rank ordering, and the basal and predict-only groups performing similarly and the lowest. This trend is evident in the rank ordering of mean proportion correct on probed recalls across all stories (predict plus justify--.75, review + justify--.73, review--.72, basal--.66, predict--.64) and the rank ordering of mean number of propositions recalled across all stories (predict plus justify--27.9, review--25.9, review plus justify--24.5, basal--23.5, predict--22.2). It should be noted that these same rank orderings were *not* evident on the pretest measures. These trends are consistent with the general expectations set forth in the original hypotheses, once the depressed performance of the predict-only group is taken into account.

Transfer Measures Analyses

Free recall. Table 9 shows the ANOVA summary table for the overall analysis of the free recall transfer measure, and Table 10 lists treatment group adjusted and observed means and standard deviations for each free recall type. In the within-subjects part of the design, there was one repeated measures factor--recall type (important literal, unimportant literal, important inferential, elaborative inferential). It can be seen that the only significant effect in the analysis, aside from the covariates, is the significant effect for recall type ($F(3,303) = 392.9, p = .000$). As in the lesson comprehension measures free recall analysis, this effect is directly interpretable and not at all surprising. Students clearly recalled more important literal information than any other type, but again, there was more of this information to recall than any other type. As Table 10 shows, the overall pattern of group differences, although not significant, follows a trend similar to the one noted throughout the lesson comprehension measures analyses--the predict-plus-justify group outperformed other groups overall. Nevertheless, this analysis of the free recall transfer task does not suggest that children in any of the experimental groups picked up any generalizable routines in the 12-day instructional period.

[Insert Tables 9 & 10 about here.]

Probed recall. Table 11 shows the ANOVA summary table and Table 12 the pattern of adjusted and observed means for the transfer probed recall task. There was one repeated measures factor in the within-subjects part of the design--question type (literal, inferential, thematic/story extension). The only significant result was a significant overall effect for question type ($F(2,202) = 16.04, p = .000$). Neither of the planned contrasts among question types--literal versus inferential or thematic versus the other two--was significant, however. Again, there is no evidence of a transfer of improved performance from an instructional to an independent reading situation in this analysis.

[Insert Tables 11 & 12 about here.]

Discussion

The most consistent finding of this study is the mediating effect that answer justification has in the case of prediction. In inference probe questions, in literal probes for one instructed story, and in overall free recall of instructed stories, the predict-plus-justify group performed significantly better than the predict-only group. Furthermore, when all five groups are rank-ordered by their performance on each outcome measure, the predict-plus-justify group nearly always ranks at the top, while the predict-only group is nearly always the lowest, or higher than only the basal group. The most compelling explanation for this discrepancy in performance between the two prediction groups is that although both were expected to engage in deep processing because of a consistent emphasis on forward-inferencing, the predict-only group's inferences may not have been sufficiently constrained by the text. The press for multiple responses to the prediction questions, although it was intended to increase reader involvement and inference making, inadvertently may have encouraged departure from the text in the predict-only group. Faced with the request to tell what might happen next in a story, and unhampered by the constraint of having to tell how they arrived at their predictions, some children may have decided that it was better to come up with a new prediction instead of agreeing with one made by someone else, in spite of the experimenter's explanation to the children that it was "OK to agree with someone else when I ask you what you think is going to happen in the story."

Early in both phases of the study, the biggest problem in both prediction groups had been the children's overconcern with being "right" in their predictions. Although stories were typed so that only one silent reading unit appeared on a page, children in the prediction groups were seen surreptitiously flipping ahead in the story, or pressing the current page down as flat as possible so they could read through to the next page, just so they could make the "right" prediction. In an effort to eradicate this over-concern with being right, the experimenter may have inadvertently pushed the balance too far in the other direction in telling the children, "Being *right* isn't the main thing. Thinking about the story

hard enough to make a prediction of what *you* think will happen is the main thing." Once correctness was no longer the major concern, one of the most common responses to the question, "And what do you think is going to happen?" was, "Well, he took mine!"--as though a prediction was invalidated by having been made already by someone else in the group.

Although an exhaustive analysis of the relationship between the content of prediction groups' discussions and their free recalls has not been made, there is some evidence that when predictions were not confirmed or refuted after further reading, they sometimes showed up as "incorrect information" in children's recalls. In the story "The Birthday," for example, in which Delilah is surprised with a puppy for her birthday even though she hinted for, and thought she was going to get, a record player, several children in each of the predict-only groups, but not in the other groups, "remembered" that she got a record player, or a puppy *and* a record player. Similarly, in another instructional story upon which lesson comprehension measures were obtained, the story's structure primes children for thinking that the main character, Stuey, will have done his homework in invisible ink the night before he actually does. Although this misconception was common in all groups before discussion, perhaps only in the predict-only groups did it go uncorrected.

It had been expected that one of the factors that would heighten prediction groups' performance was heightened engagement with the text and, thus, heightened affect. The expectation of heightened story affect in the prediction groups was disconfirmed by results from analyses of the affective response measures. Mean scores for a subset of six responses (the ones that measured suspense, excitement, surprise, interest, outcome-liking and story-liking) did not differ significantly among groups (Story 1--range = 4.7-5.3, p = .729; Story 2--range = 4.1-5.1, p = .560; Story 3--range = 3.7-5.1, p = .237), and the prediction groups were not always at the top of the range. Interestingly, when an independent rater listened to the tape recordings of the group lessons for the second story upon which lesson comprehension measures were obtained ("The Birthday"), she identified the two prediction groups as having the most lively involvement in the lesson in each phase of the study. Apparently, it is not involvement alone that increases memory for the story--rather, it may be focused engagement with text-related ideas. Predicting outcomes, justifying predictions, and confirming or refuting them upon further reading may be effective because it is in effect a whole comprehension-monitoring package in which the engagement with the text is slightly different, and more focused, at each step. Moreover, the focus of prediction on important or central text ideas in this study means that those ideas were the ones getting the increased cognitive attention in the predict-plus-justify group.

The depressed performance of the predict-only group in this study is at odds with the findings of Anderson and his associates (Anderson et al., 1987; Wilkinson et al., 1987), but it should be remembered that here, the test of prediction is much more stringent, as prediction is compared with other meaning-emphasis as opposed to word-level-emphasis conditions.

Justifying one's answers did not have the same mediating effect on performance of the review groups. Review and review-plus-justify groups performed nearly identically on many outcome measures, and neither was consistently higher than the other in cases where performance of the two groups differed. It had been expected that answer justification would provide unique opportunities for increased interaction with text ideas, clearing up misconceptions, and hearing the reasoning of others in the review-plus-justify condition. The nature of actual group interactions suggests, however, that the review-only groups had the same opportunities, but self-initiated them only as the need arose.

First, a high proportion of the review groups' discussion questions emphasized inference-making, so that text ideas already were being processed deeply. Increased interaction with the text ideas through justifying one's answers may have been just that--*increased* interaction, not necessary or better interaction. Second, recall that in review groups, more than one respondent was called on and group members were encouraged to commit themselves to an interpretation when there was disagreement. The consistent press for multiple responses in the review-only groups as well as the review-plus-justify-groups actually may have provided them with the chance for answer justification when it was most

needed--when there was disagreement about an answer. Tape recordings of review group discussions suggest that children often spontaneously justified their answers when they were disagreeing with someone else. So the review-only groups in fact may have been practicing a more economical form of answer justification and "making thinking public"--they did it when the need arose naturally, out of lack of agreement on an answer or confusion about text ideas.

The expectation that experimental groups whose story discussions focused on important and implied text ideas would outperform a basal reader control group was modestly upheld in this study. Recall that on two of the three lesson comprehension measures, performance on literal probe questions was significantly higher in experimental groups than in the basal control group; and that if the predict-only group was excluded from consideration, the effect extended to inferential probes and to the third lesson comprehension measure as well. The rank-ordering of means on the free recall lesson comprehension measures also indicates some advantage for all story map groups except the predict-only group.

Although literal information actually was the focus of more of the basal group's discussion questions than the experimental groups' questions (in fact, the two prediction groups were asked no literal questions at all), there was no systematic focus in the basal group questions on what was of central importance in the story. Thus, questions such as "What birthday was Delilah celebrating?" "What did Delilah's parents buy in the Housewares Department?" and "How would you describe Delilah's looks?" (for the story "The Birthday") proliferated in the basal story discussions, although knowing this information would not enhance performance on a test whose questions had been designed with a combined focus on important and implied text ideas. Discussion questions in the experimental groups, whether literal or inferential, focused on what was of central importance. Although probed recall questions did not duplicate any of the questions asked during experimental discussions, the same theory motivated the construction of both, and many of the same text ideas were tapped. This focus on important text ideas in story map-based group discussions also may have emphasized for these groups a set of important ideas around which to anchor their free recalls, while the basal group was left with a less systematically selected set of text ideas to use as anchor points.

The failure to find significant transfer effects was disappointing. It was encouraging, however, that the same general trends were evident for transfer free recalls and probed recalls as in the lesson comprehension measures analyses. There is some evidence in the tape-recorded lessons, in fact, that the experimental group routines may have taken some time to be learned, and that a longer study might have led to significant transfer effects. The special difficulties of the predict-only groups in learning the prediction strategy already have been noted. Similarly, early in the predict-plus-justify and review-plus-justify groups, the experimenter's request for a reason for one's answer may have been interpreted frequently as a hint that an answer was wrong and should be changed. The following example occurred in a review-plus-justify group in the first week:

T: And why did the monkey tell the crocodile he left his heart on the riverbank? (pause) What do you think? Eric?
E: Well, that was a trick on the crocodile.
T: And what makes you say that, Eric?
E: . . . I changed my mind.
T: No, I'm not saying you're wrong. I just wondered how you figured out your answer.

A more general suggestion of the time it may have taken to learn the new reading routines comes from the short strategy-related purpose-for-reading discussions at the beginning of each experimental session. In the early lessons, children were likely to focus on their behavior in the small-group lessons when asked what they were supposed to be doing in their reading groups every day. One child in a predict-plus-justify group, for example, responded, "Well, we're supposed to sit up straight, be quiet . . . and not look ahead in the story!" (The experimenter had reminded the children how important it was to listen to each other's ideas; apparently this was interpreted as "Be quiet!") By the second week of the study, there was a more general tendency at least to make an appropriate response to the strategy-

related purpose question. Interestingly, in the third week, there were even several hints of application: A child, again in a predict-plus-justify group, responded, "Like when we read the one about the invisible ink, we tried to figure out what he would do to make it (his homework) come back." Another child, this time in a review-plus-justify group, remarked that she thought it was fun to think like this and was trying it out on the stories she read at home.

Cole and Griffen (1986), in fact, argue from an educational perspective that language mediates cognitive function. Perhaps children have to learn how to talk about new ways of thinking before they can learn to do them consistently, and in a longer study more measurable treatment and transfer effects might be evident as children become more proficient as group participants.

There are several limitations in this study. First, practical constraints on assignment of children to groups meant that the assignment procedure was not truly random. Second, there is reason to believe that measurement error may have affected the results of both the lesson comprehension measure analyses and the transfer measure analyses in the ways detailed below, so that the true differences among groups must be larger than shown in this study.

First, groups differed initially in their abilities to perform the criterion tasks, especially the free recall task, in spite of the investigator's best attempts at random assignment to groups. Mean number of story propositions recalled on the pretest ranged from a low of 16.4 in the predict-only group to a high of 25.3 in the basal group. Because such large group differences existed before the intervention, it is likely that some of the variance in performance on the lesson comprehension measures that was attributed to treatment effects might have been due to students' initial abilities, even after group means were adjusted for the covariates. It is possible, then, that the basal group in particular would have had an undue advantage, and the predict-only and review groups an undue disadvantage, in the statistical analyses performed in this study. This possibility was tested by assuming a range of pretest reliabilities other than perfect reliability and studying what the results of the same analyses might have been if measurement error were taken into account. To do this, the correlations between pretest performance and each other variable in the analysis were corrected for attenuation (Lord & Novick, 1968) assuming reliabilities of .70, .80, and .90 instead of 1.0, the analyses were rerun using these corrected correlations, and the pattern of estimated results was examined.

For the lesson comprehension measure analyses, these corrections for attenuation resulted in stronger observed effects of the story map-based discussions. In the hypothetical case of pretest reliabilities of .8 and below, the story map versus basal contrast significantly favored the story map groups in both free and probed recall performance. The net effect for justifying answers also was strengthened; nevertheless, justification of answers still was especially important in the prediction condition.

Measurement error may have affected the results of the transfer analyses in an additional way. While three lesson comprehension measures were collected for each child in the study, only one transfer measure was collected. The lesson comprehension measure is thus, in effect, three times as long as the transfer measure and so can be assumed to be more reliable. One could ask how results would differ if a more reliable transfer measure (i.e., performance on more than one transfer story) were available. Again, instead of assuming perfect reliability, a range of transfer measure reliabilities was hypothesized (.70, .80, and .90), correlations between the transfer measure and all variables in each analysis were corrected for attenuation, and the original analyses were rerun. It was found that in the hypothetical case of criterion measure reliabilities of .80 or less, when corrections for attenuation were applied to the free and probed recall transfer measure analyses, the effect was to magnify the mediating effect of answer justification in the prediction condition. Furthermore, the trend was for the story map effect to carry over to independent reading. For a hypothesized reliability of .70, the story map contrast significantly favored the story map groups for free recalls, and the same trend was present (although significance was not reached) for hypothesized reliabilities of .80 and .90.

It should be emphasized, of course, that these results do no more than suggest what *might* have been found under actual conditions of initial equality of treatment groups and more reliable criterion measurement. These conditions, however, did not exist. If this study were to be replicated, several modifications such as true random assignment to treatment groups and use of several transfer measures instead of just one would certainly be warranted.

Several further explorations of these data might offer a more complete picture of exactly what may be affecting (either positively or adversely) children's ways of thinking about stories. For example, recent investigations of classroom participation structures (Au & Mason, 1981; Bloome & Green, 1984; Cazden, 1986; Green, Harker, & Golden, 1987; Weber, 1986) suggest that how children learn to think about what they read may be as much a function of how they are allowed to talk with the teacher and with one another, and how much they are allowed to initiate the questions and topics of discussion, as it is a function of the cognitive demands of teacher-posed questions. Dillon (1982, 1983) cautions that teachers' questions could do as much to constrain children's thinking as they do to promote higher intellectual functioning. In the present study, although children's questions and comments were allowed, the story discussions were driven by teacher-selected and teacher-posed questions and teacher-set rules for responding. Tape-recorded lessons from this study could be analyzed to make comparisons among groups in such things as the amount and kind of teacher talk versus student talk and the incidence of student-initiated questions or comments, because these possible differences may have influenced children's memory for stories in some way. Systematic analyses of the content of group discussions in this study, and of how the specific content of discussions showed up in or affected free and probed recalls, also have not been conducted and may be warranted. Microanalyses (e.g. Wilkinson et al., 1987) may be appropriate to explore further the extent to which unconfirmed predictions in particular may have affected story memory.

An examination of this study's results also suggests several directions for future research that may offer a more clear picture of how teachers' discussion questions affect the model of meaning that children construct for a text. For example, in the present study, students' thematic interpretations did not differ among groups (there were no group differences in performance on thematic/story extension questions and few evaluative statements in free recalls). Whether other types of teacher-posed discussion questions or other patterns of teacher-student and student-student interaction would promote these high-level responses is of future interest. Additionally, it undoubtedly is true that real, naturally written trade books for children offer far greater variety and complexity, with fewer of the constraints in length, vocabulary, and topic than those operating on the basal reader stories in this study. Elementary school teachers' discussions with children of naturally occurring trade books provides an important and almost wide-open avenue for further research (e.g., Walmsley & Walp, in press).

In any case, the present study offers important suggestions about the effect that teachers' questions can have on the model of meaning that children construct for a text. The clearest suggestion is that if teachers want to ask prediction questions about stories, they should make sure to ask children for reasons behind their predictions and provide time to monitor predictions after further reading. The study also suggests that, whether they ask prediction or review questions, teachers will do well to base their questions on text information that is important and that must be inferred. Such questions invite close attention to the text and integration of text information with prior knowledge. These forms of text processing are more desirable than processing that is unconstrained by the text (predict only) or constrained by noncentral points in the text (basal).

References

Adams, M. J., & Collins, A. (1979). A schema-theoretic view of reading. In R. O. Freedle (Ed.), *New directions in discourse processing, Vol. II: Advances in discourse processes* (pp. 1-22). Norwood, NJ: Ablex.

Anderson, R. C., & Pearson, P. D. (1984). A schema-theoretic view of basic processes in reading comprehension. In P. D. Pearson (Ed.), *Handbook of reading research* (pp. 255-291). New York: Longman.

Anderson, R. C., Wilkinson, I., Mason, J., & Shirey, L. (1987, December). *Prediction versus word-level questions*. Paper presented at the 37th meeting of the National Reading Conference, St. Petersburg Beach, FL.

Au, K. H., & Mason, J. (1981). Social organizational factors in learning to read: The balance of rights hypothesis. *Reading Research Quarterly, 17*, 115-152.

Beck, I. L., McKeown, M. G., McCaslin, E. S., & Burkes, A. M. (1979). *Instructional dimensions that may affect reading comprehension: Examples from two commercial reading programs*. Pittsburgh, PA: University of Pittsburgh Learning Research and Development Center.

Beck, I. L., Omanson, R. C., & McKeown, M. G. (1982). An instructional redesign of reading lessons: Effects on comprehension. *Reading Research Quarterly, 17*, 462-481.

Bloome, D., & Green, J. (1984). Directions in the sociolinguistic study of reading. In P. D. Pearson (Ed.), *Handbook of Reading Research* (pp. 395-421). New York: Longman.

Bruce, B. (1980a). Analysis of interacting plans as a guide of the understanding of story structure. *Poetics, 9*, 295-311.

Bruce, B. (1980b). Plans and social actions. In R. J. Spiro, B. C. Bruce, & W. F. Brewer (Eds.), *Theoretical issues in reading comprehension* (pp. 367-384). Hillsdale, NJ: Erlbaum.

Bruce, B. (1984). A new point of view on children's stories. In R. C. Anderson, J. Osborn, & R. J. Tierney (Eds.), *Learning to read in American schools: Basal readers and content texts* (pp. 153-174). Hillsdale, NJ: Erlbaum.

Bruce, B., & Newman, D. (1978). Interacting plans. *Cognitive Science, 2*, 195-233.

Cazden, C. (1986). Classroom discourse. In M. C. Wittrock (Ed.), *Handbook of research on teaching* (3rd ed., pp. 432-462). New York: Macmillan.

Cole, M., & Griffen, P. (1986). A sociohistorical approach to remediation. In S. deCastell, A. Luke, & K. Egan (Eds.), *Literacy, society, and schooling: A reader* (pp. 110-131). Cambridge, MA: Cambridge University Press.

Collins, A., Brown, J. S., & Larkin, K. M. (1980). Inference in text understanding. In R. Spiro, B. Bruce, & W. F. Brewer (Eds.), *Theoretical issues in reading comprehension* (pp. 385-407). Hillsdale, NJ: Erlbaum.

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

Dillon, J. T. (1982). The effect of questions in education and other enterprises. *Journal of Curriculum Studies, 14,* 127-152.

Dillon, J. T. (1983). *Teaching and the art of questioning* (Fastback Series No. 194). Bloomington, IN: Phi Delta Kappa.

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

Durkin, D. (1981). Reading comprehension instruction in five basal reading series. *Reading Research Quarterly, 16,* 515-544.

Fitzgerald, J. (1984). The relationship between reading ability and expectations for story structures. *Discourse Processes, 7,* 21-41.

Fitzgerald/Whaley, J. (1981). Readers' expectations for story structure. *Reading Research Quarterly, 17,* 90-114.

Fitzgerald, J., & Spiegel, D. L. (1983). Enhancing children's reading comprehension through instruction in narrative structure. *Journal of Reading Behavior, 15*(2), 1-17.

Glover, J. A., Bruning, R. H., & Plake, B. S. (1982). Distinctiveness of encoding and recall of text materials. *Journal of Educational Psychology, 74,* 522-534.

Glover, J. A., Plake, B. S., Roberts, B., Zimmer, J. W. & Palmere, M. (1981). Distinctiveness of encoding: The effects of paraphrasing and drawing inferences on memory for prose. *Journal of Educational Psychology, 73,* 736-744.

Glover, J. A., Plake, B. S., & Zimmer, J. W. (1982). Distinctiveness of encoding and memory for learning tasks. *Journal of Educational Psychology, 74,* 189-198.

Gordon, C. J., & Pearson, P. D. (1983). *The effects of instruction in metacomprehension and inferencing on children's comprehension abilities* (Tech. Rep. No. 277). Urbana-Champaign: University of Illinois, Center for the Study of Reading.

Green, J. L., Harker, J. O., & Golden, J. M. (1987). Lesson construction: Differing views. In G. W. Noblit & W. T. Pink (Eds.), *Schooling in social context: Qualitative studies* (pp. 46-77). Norwood, NJ: Ablex.

Guszak, F. (1967). Teacher questioning and reading. *The Reading Teacher, 21,* 228-234.

Hansen, J. (1981). The effects of inference training and practice on young children's reading comprehension. *Reading Research Quarterly, 16,* 391-417.

Hansen, J., & Pearson, P. D. (1983). An instructional study: Improving the inferential comprehension of good and poor fourth-grade readers. *Journal of Educational Psychology, 75,* 821-829.

Hart, C. (1977). The birthday. In E. Evertts & B. Weiss (Eds.), *Holt Basic Reading System: Level 12--Special Happenings* (pp. T77-T88 teacher's edition). New York: Holt Rinehart & Winston.

Idol, L. (1987). Group story mapping: A comprehension strategy for both skilled and unskilled readers. *Journal of Learning Disabilities, 20,* 196-205.

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

Jose, P. E., & Brewer, W. F. (1984). The development of story liking: Character identification, suspense, and outcome resolution. *Developmental Psychology, 20*, 911-924.

Lord, F. M., & Novick, M. R. (1968). *Statistical theories of mental test scores*. Menlo Park, CA: Addison-Wesley.

Mason, J., & Au, K. (1986). *Reading instruction for today*. Glenview, IL: Scott, Foresman.

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

Morrison, T. (1986). *Question types and fates of student responses in second and fourth grade basal reading instruction*. Unpublished doctoral dissertation, University of Illinois, Urbana-Champaign.

Nicholas, D. W., & Trabasso, T. (1980). Toward a taxonomy of inferences for story comprehension. In F. Wilkening, J. Becker, & T. Trabasso (Eds.). *Information integration by children* (pp. 243-265). Hillsdale, NJ: Erlbaum.

Norusis, M. J. (1986a). *Advanced statistics SPSSPC+ for the IBM PV/XT/AT*. Chicago, IL: SPSS, Inc.

Norusis, M. J. (1986b). *SPSS/PC+ for the IBM PC/XT/AT*. Chicago, IL: SPSS, Inc.

Omanson, R. C. (1982). An analysis of narratives: Identifying central, supportive, and distracting content. *Discourse Processes, 5*, 195-224.

Omanson, R. C., Beck, I. L., Voss, J. F., & McKeown, M. G. (1984). The effects of reading lessons on comprehension: A processing description. *Cognition and Instruction, 1*, 45-67.

Omanson, R. C., Warren, W. H., & Trabasso, T. (1978). Goals, inferential comprehension, and recall of stories by children. *Discourse Processes, 1*, 337-354.

Palincsar, A. S., & Brown, A. L. (1984). Reciprocal teaching of comprehension-fostering and comprehension-monitoring activities. *Cognition and Instruction, 1*, 117-175.

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

Pearson, P. D., Hansen, J., & Gordon, C. (1979). The effect of background knowledge on young children's comprehension of explicit and implicit information. *Journal of Reading Behavior, 11*, 201-209.

Pearson, P. D., & Johnson, D. (1978). *Teaching reading comprehension*. New York: Holt Rinehart & Winston.

Raphael, T., & McKinney, J. (1983). An examination of fifth- and eighth-grade children's question-answering behavior: An instructional study in metacognition. *Journal of Reading behavior, 15*(3), 67-86.

Raphael, T., & Pearson, P. D. (1985). Increasing students' awareness of sources of information for answering questions. *American Educational Research Journal, 22*, 217-236.

Raphael, T., & Wonnacott, C. A. (1985). Heightening fourth-grade students' sensitivity to sources of information for answering comprehension questions. *Reading Research Quarterly, 20*, 282-296.

Redfield, D. L., & Rousseau, E. W. (1981). A meta-analysis of experimental research on teacher questioning behavior. *Review of Educational Research, 51*, 237-245.

Rumelhart, D. (1977). Understanding and summarizing brief stories. In D. LaBerge & S. J. Samuels (Eds.), *Basic processes in reading: Perception and comprehension* (pp. 265-303). Hillsdale, NJ: Erlbaum.

Rumelhart, D. (1980). Schemata: The building blocks of cognition. In R. J. Spiro, B. C. Bruce, & W. F. Brewer (Eds.), *Theoretical issues in reading comprehension* (pp. 33-58). Hillsdale, NJ: Erlbaum.

Singer, H., & Donlan, D. (1982). Active comprehension: Problem-solving schema with question generation for comprehension of complex short stories. *Reading Research Quarterly, 17*, 166-186.

Smith, F. (1975). *Comprehension and learning.* New York: Holt Rinehart & Winston.

Stauffer, R. G. (1975). *Directing the reading-thinking process.* New York: Harper & Row.

Stein, N. L., & Glenn, C. (1979). An analysis of story comprehension in elementary school children. In R. O. Freedle (Ed.), *New directions in discourse processing, Vol. II: Advances in discourse processes* (pp. 53-120). Norwood, NJ: Ablex.

Sundbye, N. (1987). Text explicitness and inferential questioning: Effects on story understanding and recall. *Reading Research Quarterly, 22*, 82-98.

Tierney, R. J., Bridge, C., & Cerra, M. J. (1978-79). The discourse processing operations of children. *Reading Research Quarterly, 14*, 539-573.

Trabasso, T., & Nicholas, D. W. (1980). Memory and inferences in the comprehension of narratives. In F. Wilkening, J. Becker, & T. Trabasso (Eds.), *Information integration by children* (pp. 215-242). Hillsdale, NJ: Erlbaum.

Trabasso, T., Secco, T., & Van Den Broek, P. (1984). Causal cohesion and story coherence. In H. Mandl, N. L. Stein, & T. Trabasso (Eds.), *Learning and comprehension of text* (pp. 83-111). Hillsdale, NJ: Erlbaum.

Walmsley, S., & Walp, T. (in press). Toward an integrated language arts curriculum: Philosophy, practice and implications. *Elementary School Journal.*

Warren, W. H., Nicholas, D. W., & Trabasso, T. (1979). Event chains and inferences in understanding narratives. In R. O. Freedle (Ed.), *New directions in discourse processing, Vol. II: Advances in discourse processes* (pp. 23-52). Norwood, NJ: Ablex.

Weber, R. (1986, April). *The constraints of the questioning routine in reading lessons.* Paper presented at the annual meeting of the American Educational Research Association, San Francisco, CA.

Wilkinson, I., Anderson, R. C., & Wilson, P. T. (1987, December). *Prediction and the modeling of group dynamics.* Paper presented at the 37th meeting of the National Reading Conference, St. Petersburg Beach, FL.

Authors' Notes

This report is based on the first author's dissertation, which was completed at the University of Illinois at Urbana-Champaign. She gratefully acknowledges the assistance of her dissertation co-chairs Richard Anderson and P. David Pearson, dissertation committee member Robert Tierney, and Jana Mason for their advice and assistance.

The first author wishes to thank the 1985-86 principals, third-grade teachers, and third-grade students of Kenwood School in Champaign, Illinois, and Brigham and Sugar Creek Schools in Normal, Illinois, for their participation in this study.

Table 1

Planned Differences Among Groups

Group	Source of Questions	Prediction vs. Review	Justification vs. No Justification
1	Basal		
2	Story map	Review	No justification
3	Story map	Prediction	No justification
4	Story map	Review	Justification
5	Story map	Prediction	Justification

Table 2

Group Descriptive Statistics

Group	Phase	*N*	Total Reading Percentile[1]	
			M	*SD*
1-Basal	1	12	49	27
	2	8	67	27
	Total	20	56	27
2-Review	1	10	52	27
	2	12	69	18
	Total	22	61	24
3-Predict	1	12	53	25
	2	12	68	24
	Total	24	60	25
4-Review+Justify	1	13	51	30
	2	8	67	30
	Total	21	57	30
5-Predict+Justify	1	11	53	25
	2	8	68	22
	Total	19	59	25

[1]Phase 1 test = **SRA Achievement Series**, 3.1 norms, given 9/85.

Phase 2 test = **Stanford Achievement Series**, Primary 2 Form E, 3.1 norms, given 9/85.

Table 3

Story Pretest Means and Standard Deviations (*SD*) by Group

Group	Basal	Review	Predict	Review + Justify	Predict + Justify
n	20	22	24	21	19
Total Free Recall	25.3 (13.1)	18.8 (10.2)	16.4 (11.2)	16.8 (10.2)	20.7 (9.7)
Total Probed Recall	.67 (.28)	.58 (.20)	.63 (.26)	.58 (.28)	.68 (.25)

Table 4

ANOVA For Free Recall Lesson Comprehension Measures With Two Repeated Measures

Source of Variation	*SS*	*df*	*MS*	*F*	*P*
Between					
School	48.94	1	48.94	1.67	.200
Total Rdng. %tile	1418.29	1	1418.29	48.38	.000*
Pretest Tot. Recall	1883.42	1	1883.42	64.25	.000*
Contrast 1[1]	30.75	1	30.75	1.05	.309
Contrast 2	73.51	1	73.51	2.51	.117
Contrast 3	2.55	1	2.55	.09	.769
Contrast 4	191.66	1	191.66	6.54	.012*
Treat. x School	5.10	4	1.27	.04	.996
Treat. x Tot. Rdng.	21.27	4	5.32	.18	.947
Treat. x Pretest	93.39	4	23.35	.80	.531
Residual	2521.08	86	29.31		
Within					
Recall Type	44795.86	3	14931.95	714.82	.000*
Treat. x Recall Type	298.75	12	24.90	1.19	.288
Residual	6329.42	303	20.89		
Story	939.77	2	469.88	35.75	.000*
Treat. x Story	141.85	8	17.73	1.35	.221
Residual	2654.92	202	13.14		
Recall Type x Story	7642.12	6	1273.69	172.47	.000*
Treat. x Rec. Type x Story	236.97	24	9.87	1.34	.131
Residual	4475.21	606	7.38		

[1]Contrast 1: Story map
Contrast 2: Justification
Contrast 3: Question focus (review vs. predict)
Contrast 4: Justification by question focus

Table 5

Mean Total Free Recall Across All Stories by Group

Group	Basal	Review	Predict	Review + Justify	Predict + Justify
M					
Adjusted	23.4	25.8	22.2	24.5	27.9
Unadjusted	24.0	25.3	20.7	23.3	28.6
(*SD*)	11.8	9.1	9.1	11.2	10.9

Table 6

ANOVA For Probed Recall Interim Measures With Two Repeated Measures

Source of Variation	*SS*	*df*	*MS*	*F*	*P*
Between					
School	.44	1	.44	3.73	.056
Total Rdng. %tile	6.02	1	6.02	51.00	.000*
Pretest Total	3.69	1	3.69	31.27	.000*
Contrast 1	.29	1	.29	2.43	.123
Contrast 2	.74	1	.74	6.25	.014*
Contrast 3	.18	1	.18	1.49	.225
Treatment x School	.81	4	.20	1.72	.155
Treat. x Tot. Rdng.	.31	4	.08	.66	.628
Treat. x Pretest	.55	4	.14	1.16	.332
Residual	10.16	86	.118		
Within					
Question type	3.91	2	1.95	54.02	.000*
Treat. x Q. type	.65	8	.08	2.24	.026*
Residual	7.31	202	.036		
Story	.08	2	.04	.91	.406
Treat. x Story	.33	8	.04	.94	.486
Residual	8.88	202	.044		
Q. type x Story	3.12	4	.78	30.62	.000*
Treat x Q. type x Story	.74	16	.05	1.82	.026*
Residual	10.28	404	.025		

Table 7

Adjusted and Observed Means and Standard Deviations (*SD*) for Lesson Comprehension Probed Recall Question Types

Group	Literal			Inferential			Thematic/ Story Extension			Total Probed Recall		
Story 1	adj.	obs.	(*SD*)	adj.	obs.	(*SD*)	adj.	obs.	(*SD*)	adj.	obs.	(*SD*)
Basal	.62	.62	(.28)	.75	.75	(.27)	.60	.60	(.28)	.65	.66	(.25)
Review	.70	.69	(.16)	.78	.78	(.24)	.62	.61	(.28)	.70	.69	(.20)
Predict	.61	.62	(.14)	.66	.66	(.25)	.57	.58	(.22)	.61	.62	(.18)
Review + Justify	.74	.72	(.20)	.78	.76	(.27)	.62	.60	(.19)	.72	.71	(.16)
Predict + Justify	.77	.78	(.22)	.84	.86	(.23)	.66	.68	(.22)	.76	.77	(.18)
Story 1 Total		.68	(.21)		.76	(.25)		.61	(.24)		.68	(.24)
Story 2												
Basal	.69	.69	(.27)	.56	.56	(.27)	.68	.68	(.27)	.65	.65	(.21)
Review	.83	.83	(.14)	.70	.69	(.24)	.63	.63	(.22)	.72	.72	(.15)
Predict	.81	.82	(.17)	.48	.48	(.15)	.70	.70	(.27)	.66	.67	(.14)
Review + Justify	.84	.82	(.21)	.71	.69	(.24)	.76	.75	(.24)	.77	.75	(.19)

Table 7 (continued)

Group	Literal			Inferential			Thematic/ Story Extension			Total Probed Recall		
Story 2	adj.	obs.	(*SD*)	adj.	obs.	(*SD*)	adj.	obs.	(*SD*)	adj.	obs.	(*SD*)
Predict + Justify	.92	.92	(.11)	.66	.67	(.17)	.67	.68	(.22)	.75	.76	(.11)
Story 2 Total		.82	(.20)		.62	(.23)		.69	(.24)		.71	(.22)
Story 3												
Basal	.84	.85	(.14)	.64	.66	(.29)	.57	.58	(.32)	.69	.69	(.21)
Review	.89	.88	(.11)	.74	.72	(.26)	.58	.57	(.20)	.74	.73	(.15)
Predict	.81	.82	(.23)	.58	.59	(.29)	.54	.54	(.25)	.65	.65	(.22)
Review + Justify	.87	.84	(.22)	.70	.67	(.27)	.54	.52	(.30)	.70	.68	(.23)
Predict + Justify	.80	.82	(.19)	.81	.83	(.19)	.65	.67	(.21)	.75	.77	(.13)
Story 3 Total		.84	(.18)		.69	(.25)		.65	(.25)		.71	(.24)

Table 8

Group Differences For Each Combination of Story and Question Type

Contrast	1		2		3		4	
	Basal vs. Others		Justify vs. No Justify		Review vs. Predict		Interaction (C2 x C3)	
Story 1	*F**	*p*	*F*	*p*	*F*	*p*	*F*	*p*
Literal	4.07	.046*	6.78	.011*	.71	.403	2.28	.134
Inferential	.10	.755	3.85	.053	.40	.530	3.21	.076
Thematic/ Story Ext.	.14	.713	.95	.333	.00	.961	.97	.328
Story 2								
Literal	12.29	.001*	2.27	.135	.69	.409	1.48	.227
Inferential	2.49	.118	5.13	.026*	11.09	.001*	4.35	.040*
Thematic/ Story Ext.	.01	.921	1.04	.311	.04	.836	2.31	.131
Story 3								
Literal	.00	.990	.17	.680	3.83	.055	.01	.932
Inferential	1.49	.225	3.37	.070	.32	.571	7.55	.007*
Thematic/ Story Ext.	.00	.952	.43	.513	.40	.528	2.01	.159

*Critical $F(5.86) = 6.98$ for $p = .01$

Critical $F(5,86) = 3.96$ for $p = .05$

Table 9

ANOVA For Free Recall Transfer Measure

Source of Variation	*SS*	*df*	*MS*	*F*	*P*
Between					
School	234.29	1	234.29	6.71	.011*
Total Reading %tile	1978.91	1	1978.91	56.65	.000*
Pretest Total	1359.44	1	1359.44	38.91	.000*
Contrast 1	31.47	1	31.47	.90	.345
Contrast 2	.85	1	.85	.02	.877
Contrast 3	.10	1	.10	.00	.958
Contrast 4	106.29	1	106.29	3.04	.085
Treatment x School	110.95	4	27.74	.79	.432
Treat. x Tot. Rdng.	136.30	4	34.07	.98	.425
Treat. x Pretest	73.73	4	18.43	.53	.716
Residual	3004.40	86			
Within					
Recall Type	32771.08	3	10923.70	392.90	.000*
Treat. x Rec. Type	270.97	12	22.58	.81	.638
Residual	8424.13	303	27.8		

Table 10

Adjusted and Observed Means and Standard Deviations (*SD*) For Transfer Free Recall

Group	Basal	Review	Predict	Review + Justify	Predict + Justify
Important Literal					
adj.	22.9	26.2	22.8	23.6	25.1
obs.	25.1	26.4	21.7	21.9	25.7
SD	(11.0)	(8.6)	(11.8)	(11.9)	(12.5)
Unimportant Literal					
adj.	4.2	5.6	4.4	4.1	5.8
obs.	4.8	5.5	4.0	3.7	5.9
SD	(3.2)	(3.2)	(3.3)	(3.0)	(4.9)
Important Inferential					
adj.	2.8	3.3	3.8	3.3	3.4
obs.	3.1	3.3	3.7	3.1	3.5
SD	(2.3)	(2.2)	(2.6)	(2.5)	(2.7)
Elaborative Inferential					
adj.	3.3	3.0	2.8	3.0	4.4
obs.	3.5	3.0	2.6	2.8	4.4
SD	(2.1)	(2.1)	(2.0)	(2.5)	(3.3)
Overall					
adj.	33.3	38.0	33.7	34.0	38.7
obs.	36.5	38.2	32.0	31.5	39.5
SD	(16.2)	(12.5)	(16.6)	(17.4)	(18.7)

Table 11

ANOVA For Probed Recall Transfer Measure

Source of Variation	*SS*	*df*	*MS*	*F*	*P*
Between					
School	.69	1	.69	5.95	.017*
Total Reading %tile	5.17	1	5.17	44.57	.001*
Pretest Total	2.89	1	2.89	24.91	.001*
Contrast 1	.0	1	.0	.0	.934
Contrast 2	.01	1	.01	.11	.728
Contrast 3	.01	1	.01	.11	.735
Contrast 4	.28	1	.28	2.44	.122
Treatment x School	.63	4	.16	1.36	.257
Treat. x Tot. Rdng.	.41	4	.10	.88	.482
Treat. x Pretest	.44	4	.11	.94	.448
Residual	10.02	86	.116		
Within					
Question Type	.77	2	3.85	16.04	.000*
Treat. x Ques.	.14	8	.017	.74	.652
Residual	4.76	202	.024		

Table 12

Adjusted and Observed Means and Standard Deviations (*SD*) For Transfer Probed Recall

Group	Basal	Review	Predict	Review + Justify	Predict + Justify
Literal					
adj.	.78	.83	.79	.83	.87
obs.	.78	.83	.80	.80	.90
SD	(.31)	(.25)	(.25)	(.31)	(.21)
Inferential					
adj.	.76	.81	.68	.70	.79
obs.	.77	.80	.69	.67	.82
SD	(.32)	(.25)	(.30)	(.28)	(.21)
Thematic/ Story Extension					
adj.	.71	.72	.67	.68	.72
obs.	.72	.71	.67	.65	.74
SD	(.35)	(.26)	(.32)	(.34)	(.28)
Total					
adj.	.75	.79	.71	.74	.79
obs.	.76	.78	.72	.71	.82
SD	(.30)	(.22)	(.27)	(.28)	(.20)

This page is intentionally blank.