Learning and Individual Differences 22 (2012) 100–111

Contents lists available at SciVerse ScienceDirect

Learning and Individual Differences

journal homepage: www.elsevier.com/locate/lindif

Making the right connections: Differential effects of reading intervention for subgroups of comprehenders

Kristen L. McMaster [a,*], Paul van den Broek [b], Christine A. Espin [b], Mary Jane White [a], David N. Rapp [c], Panayiota Kendeou [d], Catherine M. Bohn-Gettler [e], Sarah Carlson [a]

[a] University of Minnesota, United States
[b] University of Leiden, Netherlands
[c] Northwestern University, United States
[d] Neapolis University Pafos, Cyprus
[e] Wichita State University, United States

ARTICLE INFO

Article history:
Received 14 April 2011
Received in revised form 29 October 2011
Accepted 24 November 2011

Keywords:
Reading intervention
Reading comprehension
Struggling readers
Cognitive profiles

ABSTRACT

The purpose of this study was to compare the effects of different types of questioning interventions on students' reading comprehension. Fourth-grade students ($n = 246$) were identified as struggling, average, or good readers and assigned randomly within school to one of three questioning interventions: two inferential conditions (Causal or General) or one literal condition ("Who, What, Where, When" or W-questioning). Teachers delivered the interventions for 20–30 min, 2–4 times per week, for 8–10 weeks. All readers made reliable pre- to posttest comprehension gains as measured by story recall ($ps < .001$ to .04). Differential effects for intervention were found between two subgroups of struggling comprehenders—elaborators and paraphrasers. Elaborators benefited more than paraphrasers from Causal questioning ($d = .86$) whereas paraphrasers benefited more than elaborators from General questioning ($d = 1.46$). These findings suggest that identifying subgroups is important in developing and evaluating the effectiveness of reading comprehension interventions.

1. Introduction

The failure of many students to become proficient readers is a persistent problem. According to the 2009 Nation's Report Card (National Center for Education Statistics, 2010), 33% of U.S. fourth graders and 25% of eighth graders read below basic levels. Students who experience reading difficulties are likely to struggle throughout school and to experience difficulties with employment and other aspects of daily living as adults (Snow, 2002).

As students advance through school, they must be able to read and understand a wide range of content-related materials (Chall, 1996). Some students begin to struggle when academic requirements shift to include comprehension of increasingly complex texts. Thus, a substantial amount of research has focused on the development of reading comprehension interventions.

Recent research syntheses (e.g., Berkeley et al., 2010; Gersten et al., 2001; National Institute of Child Health and Human Development, 2000) point to a wide array of interventions that have improved comprehension outcomes for struggling readers. Yet, despite the extensive knowledge base on effective reading comprehension interventions, the gap between struggling readers and their peers becomes larger and more difficult to close as they advance through school (Faggella-Luby & Deshler, 2008). Attempts to close this gap might be aided by understanding *under what instructional conditions* and *for whom* specific interventions are effective (Connor et al., 2004).

The purpose of the present study was to explore the effectiveness of specific interventions and determine whether intervention effectiveness varies with reader characteristics. To this end, we draw upon cognitive theories of reading comprehension, following recommendations to link cognitive theory and educational practice to develop more precise and focused interventions (e.g., McKeown et al., 2009; Pressley et al., 2006; Rapp, van den Broek, McMaster, Kendeou, & Espin, 2007).

1.1. Connecting cognitive science to reading intervention

Cognitive scientists emphasize that successful reading comprehension depends on the construction of a coherent representation of text in memory (Kintsch, 1998; van den Broek, 2010). A coherent text representation is formed when information in the text is integrated with the reader's background knowledge. This coherent representation is easily accessible and can be applied in a variety of situations (e.g., Goldman & Varnhagen, 1986; Graesser & Clark, 1985; Kintsch & van Dijk, 1978; Trabasso et al., 1984). Prompts designed to promote inference generation can be used to help readers construct a coherent text representation. Such prompts guide the reader to connect information within

* Corresponding author at: University of Minnesota, 250 Education Sciences Building, 56 East River Road, Minneapolis, MN 55455, United States.
E-mail address: mcmas004@umn.edu (K.L. McMaster).

1041-6080/$ – see front matter © 2011 Elsevier Inc. All rights reserved.
doi:10.1016/j.lindif.2011.11.017

various parts of the text, and to connect information in the text to relevant background knowledge.

Considerable evidence indicates that explicit inference instruction, questioning approaches, or a combination of both can have positive effects on the reading comprehension of struggling readers (e.g., Hansen & Pearson, 1983; McGee & Johnson, 2003; Pressley et al., 1987; Yuill & Joscelyne, 1988; Yuill & Oakhill, 1988). Further, researchers have provided evidence that prompting inferences *during* reading might be more beneficial than after reading (Laing & Kamhi, 2002; van den Broek et al., 2001), presumably because it is during the moment-by-moment reading process that readers construct coherent representations of text (e.g., Goldman & Varma, 1995; Kintsch, 1998; van den Broek et al., 1998).

Cognitive scientists have also identified specific *types* of connections that are fundamental to the reader's construction of coherent representations of text. Specifically, identifying *causal relations* among events or facts within a text is critical to comprehension (Trabasso & van den Broek, 1985). To illustrate, in the following text: "The lady gave the waiter $20. He returned to give her the change," the reader establishes causal coherence—and thus comprehends the text—by inferring that the waiter is returning change because the lady's lunch cost less than $20. The causal relations that readers must infer are usually more complex than in this example, extending over long distances in the text and requiring coordination of multiple pieces of information (van den Broek et al., 2009).

Given that establishing causal relations is important for coherence, questions designed to prompt such causal connections should increase the impact of interventions focused on improving reading comprehension. In the example above, a question that prompts a causal connection, such as "Why did the waiter return to give the lady the change?" should lead the reader to infer that the lady's lunch cost less than $20. This type of question should direct the reader to attend to specific causal information in the text. In a recent study (van den Broek, McMaster, Rapp, Kendeou, Espin, & Deno, 2006), we compared causal questioning approaches to general questions ("How does this sentence relate to something you previously learned in the text?") in brief, one-on-one experimental sessions. We found that, when asked causal questions, readers recalled more causally-connected events in narrative text than they recalled when asked general questions.

Evidence also suggests that direct prompting of inference generation might be more effective than explicitly teaching comprehension strategies to promote students' text comprehension. McKeown et al. (2009) compared the effects of two reading interventions designed to promote fifth graders' comprehension by prompting text-based connections. Fifth-grade classes were assigned randomly to a text-processing intervention or to strategies instruction. In the text-processing approach, the teacher read with the students, stopping at key points in the text to ask questions that prompted text connections. In the strategies approach, the teacher also stopped at key points in the text, but taught students to use strategies (e.g., summarizing, comprehension monitoring) to understand the text. McKeown et al. found that the text-processing approach had a stronger effect on students' comprehension than did the strategies approach. They noted that they did not examine how results varied by student reading level, and thus they encouraged future intervention-by-reader-type examinations.

1.2. Subgroups of struggling readers

As suggested by McKeown et al. (2009), in addition to establishing effective approaches to helping students make important text-based connections during reading, it is important to determine *for whom* different approaches are most effective (Connor et al., 2004; Faggella-Luby & Deshler, 2008). Researchers have found that children who struggle with reading comprehension do not fit one specific profile (e.g., Cain & Oakhill, 2006; Nation et al., 2002), but instead show varying patterns of weak verbal or cognitive skills that affect their reading development in different ways. If struggling readers have different types of comprehension difficulties, they might also respond to interventions in different ways.

In a recent series of studies (see Rapp et al., 2007), we observed that struggling readers have different types of inferencing difficulties. We asked fourth-grade students to think aloud as they read narrative text, and coded their responses into 11 possible categories: (1) associations to background knowledge, (2) connecting inferences, (3) reinstatement inferences, (4) elaborative inferences, (5) predictive inferences, (6) comprehension monitoring, (7) paraphrases, (8) repetitions of the text, (9) affective statements (10) evaluations of the text, and (11) questions about the text. Using cluster analysis, we identified different sub-groups of struggling readers based on their think-aloud responses. The 11 types of responses were treated as distinct cases in this analysis. We used Ward's method (Ward & Hook, 1963), which attempts to minimize the sum of squares of observations within any two clusters that are formed at each step. Three different clustering solutions were considered (i.e., two-, three-, and four-cluster solutions). The two-cluster solution was adopted as the best description of the dataset because the three- and four-cluster solutions did not account for significant additional variability in the data. The identification of two subgroups of struggling readers was replicated in three separate samples over three years of research (Rapp et al., 2007; van den Broek et al., 2006).

The two subgroups of struggling readers differed in the types of inferences they made during reading. The first subgroup, which we call 'elaborators', made the same number of elaborative inferences as did average and good readers, but their inferences were more likely to be inaccurate or invalid. The second subgroup, which we call 'paraphrasers', made fewer inferences than did the average and good readers (in fact, made few inferences in general), and instead repeated or paraphrased the text. Support for the first subgroup is provided by research that suggests that some struggling comprehenders have difficulty constructing coherent representations of text due to inappropriate use of background knowledge or personal viewpoints (e.g., Williams, 1993). Support for the second subgroup is provided by research that demonstrates that some struggling comprehenders fail to generate many inferences and thus have difficulty establishing coherence (e.g., Cain & Oakhill, 2006).

Elaborators and paraphrasers did not differ on standardized measures of listening or reading comprehension, oral reading fluency, decoding and word recognition, vocabulary, general intelligence, motivation, or working memory (Rapp et al., 2007). These findings support the view that think-aloud tasks, designed to measure reading *processes*, may yield important, unique information about struggling readers that goes undetected by more traditional measures designed to assess the *products* of reading (e.g., Ericsson & Simon, 1993; Kendeou et al., 2010; Pressley & Afflerbach, 1995). An alternative explanation, however, might be that the subgroups are merely an artifact of the think-aloud procedure. An important question, then, is whether the subgroups are instructionally relevant; that is, does identification of subgroups have implications for how to provide instruction to struggling readers? A strong test of the instructional validity of these subgroups would be to examine whether elaborators and paraphrasers respond differently to different types of questioning interventions.

1.3. Study purpose and research questions

The literature reviewed above provides important directions for further research. First, asking questions to prompt readers to make inferences during reading appears to be a promising approach to improving comprehension of text (e.g., Hansen & Pearson, 1983; McGee & Johnson, 2003; McKeown et al., 2009; Pressley et al., 1987; van den Broek et al., 2001; Yuill & Joscelyne, 1988; Yuill & Oakhill, 1988). Second, given that establishing causal relations is critical to coherence (Trabasso & van den Broek, 1985), questions designed to prompt causal connections might increase the effect of interventions

focused on improving reading comprehension (Rapp et al., 2007). Third, research is needed to determine whether questioning approaches designed to enhance readers' comprehension are differentially effective for different types of readers (cf. McKeown et al., 2009).

In the present study, we addressed the above issues in the following ways. First, we compared two questioning approaches designed to prompt inference generation (Causal questioning vs. General questioning) to a literal questioning approach ("Who, What, Where, When" or W-questioning). We expected that the inferential approaches (Causal and General) would be more effective than the literal questioning approach, because inferential questions should prompt readers to establish important connections during reading. Further, we expected that Causal questions would prompt readers to make more *relevant* connections within text than would General questions, because identifying causal relations is necessary for establishing a coherent representation of text (Trabasso & van den Broek, 1985).

Second, we examined whether different types of readers (struggling, average, good) and subgroups of struggling readers (elaborators and paraphrasers) responded differently to Causal, General, and W-questioning approaches. By comparing the responses of elaborators and paraphrasers to different types of questioning techniques, we can test the instructional relevance of subgroups of struggling readers. We expected Causal questioning to be relatively more beneficial to elaborators, because Causal questions should prompt relevant connections specific to the text (recall that in our previous research, elaborators made many connections, but their connections were often irrelevant or inaccurate). In contrast, we expected General questioning to be relatively more beneficial for paraphrasers, because General questions should prompt connections beyond the immediate sentence (recall that in our previous research, paraphrasers did not make connections to either background knowledge or to the text).

To address these two purposes, the following research questions were addressed: (1) What is the effect of Causal vs. General vs. W-questioning on the text recall of struggling, average, and good readers? (2) Is there a differential effect of Causal vs. General vs. W-questioning on the text recall of two subgroups of struggling readers (elaborators and paraphrasers)?

2. Method

2.1. Research design and study overview

For Research Question 1, we used a pretest/posttest design with random assignment to intervention groups. For Research Question 2, we used a quasi-experimental pretest/posttest design. A timeline of study activities is shown in Fig. 1. Students were first screened to identify struggling, average, and good readers (as described below) and grouped by "instructional team," defined as a group of participating classrooms in one school building that received language arts instruction at the same time of day. Students were stratified based on struggling, average, or good reader status and assigned randomly to one of three questioning interventions: Causal, General, or W-questioning (which served as the control). Students were pretested and given a think-aloud task to identify elaborators and paraphrasers. Students were then given nine weeks of classroom-based intervention.

Each intervention was implemented with at least one intervention group in each school. Intervention group sizes ranged from 23 to 28 students (which is a typical fourth-grade class size in the schools in which the study was conducted). To minimize teacher effects, students remained in the same intervention group throughout the study, and teachers rotated to a new intervention group at 3- to 4-week intervals. Based on the number of classrooms, there were 13 intervention groups; based on random assignment, there were 4 groups each that received General and W-questioning and 5 groups that received Causal Questioning. After posttesting, analyses were conducted comparing struggling, average, and good readers' responses to the three questioning interventions, and comparing elaborators' and paraphrasers' responses to the three questioning interventions.

2.2. Setting and participants

2.2.1. Schools

Participants were teachers and students from three elementary schools in two first-ring suburbs of a large, upper Midwest, urban area. The first district, from which one school participated, served 4540 students; 77% were White, 7% African American, 6% Hispanic, 8% Asian, and 2% American Indian. Twenty-seven percent of the students received free or reduced-price lunch, 11% qualified for special education, and 6% qualified for English Language Learner (ELL) services. The second district, from which two schools participated, served 9063 students; 66% were White, 14% African American, 10% Hispanic, 9% Asian, and 1% American Indian. Twenty-nine percent of the students received free or reduced-price lunch, 12% qualified for special education, and 10% qualified for ELL services.

2.2.2. Teachers

Thirteen fourth-grade teachers (4 male) participated. Teachers ranged in age from 30 to 59 years old, had an average of 21.8 years of teaching experience (range: 4 to 38 years), and had taught in their current positions for an average of 11 years (range: 2 to 28 years). Each teacher implemented each type of questioning intervention, rotating to a new intervention group at 3- to 4-week intervals, to avoid the confound of teacher effects.

Timeline of Study Activities

	FALL			FALL/WINTER				WINTER/SPRING
RQ1	Screening to identify struggling, average, & good readers	→	Stratified random assignment to Causal, General, or W-Questioning	→ Pretest	→ 9-week intervention	→ Posttest		Comparison of struggling, average, and good readers' response to 3 questioning conditions
RQ2				Think-aloud task and coding		Cluster analysis to identify elaborators and paraphrasers		Comparison of elaborators' and paraphraser's response to 3 questioning conditions

Note. RQ1 = Research Question 1:What is the effect of Causal vs. General vs. W-Questioning on the text recall of struggling, average, and good readers?
RQ 2 = Research Question 2: Is there a differential effect of Causal vs. General vs. W-Questioning on thetext recall of two subgroups of struggling readers (elaborators and paraphrasers)?

Fig. 1. Timeline of study activities.

2.2.3. Students

A total of 246 fourth-grade students (123 male) participated. The students' mean age was 9 years, 5 months; 34 were African American, 7 Asian, 199 Caucasian, 14 Hispanic, and 11 identified as "Other." All spoke English as their native language. Students with developmental or intellectual disabilities and ELLs with very limited English proficiency were not included in the study. The original sample was 265 students (134 male), but 19 students dropped from the study because of incomplete data due to moving or excessive absences. Students were initially screened for the study using a Curriculum-Based Measurement (CBM) Maze task (Fuchs & Fuchs, 1992) and the Gates–MacGinitie Reading Test comprehension subtest (MacGinitie et al., 2002; see Measures). There were no statistically significant differences between students who dropped from the study and those who remained with respect to Maze ($F[1, 261] = 1.058$, $p = .305$) or GMRT ($F[1, 261] = .012$, $p = .914$).

Students were placed into struggling, average, and good reader groups based on CBM Maze scores, and verified by GMRT scores. Distributions of Maze scores were used to determine cut-points that were based on previous research (Rapp et al., 2007; van den Broek et al., 2006). Struggling readers ($n = 56$) scored at the 23rd percentile (at or below 3.5 correct maze choices; CMC). Average readers ($n = 129$) scored between the 26th and 74th percentile (3.51 to 8.00 CMC). Good readers ($n = 59$) scored above the 74th percentile (at or above 8.17 CMC). Multivariate analyses of variance (MANOVAs) with group as the independent variable and Maze and GMRT scores as dependent variables confirmed differences among the three groups (see Table 1 for means, *SD*s, and *F*-values). Post-hoc analyses using the Least Significant Difference (LSD) indicated that struggling readers performed significantly lower than average and good readers, and average readers performed significantly lower than good readers (all $ps < .001$).

2.3. Measures and materials

2.3.1. Screening

The CBM Maze task (Fuchs & Fuchs, 1992) consisted of three passages in which every seventh word was deleted and replaced with three word choices, only one of which fit in the sentence. Students completed two practice sentences, and then read each passage silently for 1 min, selecting the words that best fit in the sentences. The score was the average CMC across all three texts. Total administration time for the three texts ranged from 5 to 10 min. Correlations between maze selection scores and other reading measures typically range from $r = .60$ to .80 for elementary-aged students (Wayman et al., 2007).

The GMRT Comprehension subtest (MacGinitie et al., 2002) includes 11 passages and 48 multiple choice questions. Students were given instructions and a practice text with questions. Then, they independently read passages and answered questions, recording answers on a separate sheet. The test was completed in about 45 min. The raw score was the number of items answered correctly. MacGinitie et al. (2002) reported reliability for fourth graders as $r = .93$; content validity was established through examination of completion rates, floor and ceiling data, and the use of inferential and literal questions throughout the test.

Table 1
Means, SDs, and F-values for CBM Maze and GMRT screening measures.

	Struggling ($n = 56$)		Average ($n = 129$)		Good ($n = 59$)		$F(2, 241)$
	M	*SD*	*M*	*SD*	*M*	*SD*	
Screening measure							
CBM Maze	2.30	(1.26)	6.61	(3.25)	11.24	(3.79)	122.33*
GMRT	31.27	(15.98)	51.09	(16.70)	69.17	(15.27)	78.64*

Note. GRMT = Gates MacGinitie Reading Test.
* $p < .001$.

2.3.2. Pre- and posttesting

Students' recall of narrative texts provided dependent variables at pre- and posttest. Pre- and posttests texts were developed by our research team.

2.3.2.1. Text development. Two pairs of narrative texts were created for pre- and posttesting. Texts within each pair were similar in terms of plot, goal structure, the general nature of events, and the tense of the language, but differed in terms of characters, setting, and details pertaining to events. Each text had a goal structure with a main goal that spanned the length of the text situation and subgoals related to the main goal. Once the general goal structure was developed and events were included to expand on this goal structure, the texts were entered into Coh-Metrix (Graesser et al., 2004), a computer tool that analyzes texts for cohesion, language (such as parts of speech, word frequency, and semantics), and readability. Texts were reworded as needed to increase cohesion while keeping the number of words, sentences, and readability levels constant (Flesch–Kincaid grade levels = 4.00 to 4.65). Each text was printed, single-spaced, on 8 in. × 11 in. card stock paper in size 14 Times New Roman font.

After the texts were developed, the events in each text were divided into idea units that included a subject and a verb. To establish a causal network for each text, each idea unit was connected to other events in the story that either caused or enabled it, following procedures by Trabasso and colleagues (e.g., Trabasso & van den Broek, 1985). We used information from the causal analyses to code students' recall of each story.

2.3.2.2. Administration. Each student was assigned to read two texts for recall at pre-test and the remaining two at post-test. Text order was counterbalanced across students, such that each student read one text from each pair at pretest and the other text from each pair at posttest, in random order. The examiner gave each student a practice task in which the student was asked to read a simple narrative aloud, then to retell the story without referring to the text. After the practice task, the examiner presented the two texts, one at a time. For each text, the student read the passage aloud and did not receive decoding help. The examiner then set the story aside and asked the student to retell the story. The entire session was audio recorded.

To confirm that text order did not influence student outcomes across intervention groups, we conducted repeated measures multivariate analyses of variance (RM-MANOVA) with text order and intervention condition as between-subjects factors, time as the within-subjects factor, and different types of items recalled (described below under "Coding") as outcome variables. There were no statistically significant interactions among time, intervention condition, or text order at pretest, Wilks' Lambda = .986, $F(7, 219) = .45$, $p = .871$; or at posttest, Wilks' Lambda = .979, $F(7, 219) = .672$, $p = .979$. Thus, text effects did not confound study outcomes.

2.3.2.3. Coding of recalls. Student recalls of each text were transcribed and parsed. Each parsed clause was compared to the parsing of the original narrative and coded according to the event that it most closely matched in the corresponding narrative. Further, each clause was categorized based on how closely it matched to the gist of the original text. This categorization was adapted from a more refined, inferential coding scheme (see Kendeou & van den Broek, 2005; Linderholm & van den Broek, 2002), and included five categories, described below (for specific examples of coded responses, see Appendix A). Categories 1 to 3 concern clauses in the recall protocols that could be directly matched to story units within the text:

1. *Conservative*: The clause in the recall protocol is a (nearly) literal rendering of the original text unit. Recall of a high proportion of conservative clauses reflects (near-)verbatim memory for the text.

2. *Liberal*: The clause in the recall protocol is a non-literal rendering of the original text unit, capturing the essence of its meaning. Recall of a high proportion of liberal clauses reflects the extent to which the reader has established a coherent representation of the text.
3. *Highly connected*: The clause in the recall protocol has five or more causal connections based on the causal network for that text. These clauses are central to the structure of the text. Recall of a high proportion of these clauses indicates the reader's strong sensitivity to the text structure. Highly connected items were also coded as conservative or liberal, based on the definitions above.

Categories 4 and 5 represent recall of clauses that could *not* be matched directly with the text:

4. *No match consistent*: The clause in the recall protocol could not be matched directly with the gist of a text unit; however, the clause was valid and moderately constrained by the text. Inclusion of these clauses in recall reflects the extent to which the reader has elaborated his/her mental representation through inferences.
5. *No match inconsistent*: The clause in the recall protocol does not match directly with the gist of a text unit and was invalid or unconstrained by the text. A high proportion of no match consistent units recalled suggests that the reader has constructed an erroneous, or partly erroneous, representation of the text.

For analyses, we used the number of unique conservative, liberal, highly-connected, no match consistent, or no match inconsistent units recalled in the two stories (administered at both pre- and post-test), divided by the total number of story units in the two stories. Note that the proportion of items recalled by each reader does not add up to 100% for three reasons. First, we only counted 'unique' items recalled. 'Unique' items represent one idea in the text (i.e., we did not count repeated recalls of one idea more than once). Second, a reader may not have recalled all of the units in the story, or may have recalled ideas beyond those found directly in the text (i.e., no match consistent and inconsistent recalls). Third, 'highly connected' items were a subset of the items that were coded as conservative or liberal, as they were those items that had five or more causal connections based on the causal network of the text.

2.3.3. Identification of subgroups of struggling readers

As in previous related studies (Authors, 2006, Rapp et al., 2007; van den Broek et al., 2006), we used a think-aloud task to identify subgroups of struggling readers. This task was presented following pretesting.

2.3.3.1. Text development. The think-aloud task was adapted from Ericsson and Simon (1993) for use in this study. Each sentence from each text was typed on 5 in. × 6 in. cards. The practice text ("Jimmy and the New Bike," adapted from Magliano et al., 1999) and test text ("Brian's Magical Skill," adapted from van den Broek et al., 2001) are provided in Appendix B.

2.3.3.2. Administration. The examiner demonstrated the think-aloud task with the first half of the practice text by giving specific examples of all possible types of responses (e.g., elaborative inferences, paraphrases, predictions). Then, the student practiced thinking aloud with the remaining sentences in the practice text. The student read one sentence at a time and then told the examiner what he or she was thinking. After each text, the examiner asked the student two yes/no questions about the story to check for general comprehension. The student received no decoding help, but received non-leading prompts such as, "What are you thinking now?" if he or she forgot to think-aloud. The entire session was audio recorded.

2.3.3.3. Coding of think-alouds. Each clause in the think-aloud task was categorized based on a coding scheme adapted from previous research (Linderholm & van den Broek, 2002) designed to identify the cognitive processes in which readers engaged during reading. The primary variables coded included: *elaborative inferences* (readers' attempts to explain the contents of the current sentence using background knowledge, coded as *valid* when explanations were constrained by the text and error free, and *invalid* when they were not constrained by the text or involved error), *predictive inferences* (readers' anticipation of what will occur next in the text, also coded as *valid* or *invalid* based on whether they were constrained by the text or involved error), *connecting inferences* (readers' explanation of the current sentence by connecting it to the immediately preceding sentence), *reinstatement inferences* (readers' explanation of the current sentence on the basis of prior text information not in the immediately preceding sentence), *metacognitive responses* (readers' reflections on their understanding of text), *evaluations* (readers' opinions about the text), and *paraphrases/text repetitions* (readers' restatement of the current sentence in their own words or verbatim reiterations).

2.4. Questioning interventions

2.4.1. Intervention structure

All students participated in the intervention as part of regular classroom instruction. Teachers were asked to implement the intervention 3 times per week for 20 to 30 min per session across 9 weeks, for a total of 15 h. On rare occasions, because of disruptions to school schedules (e.g., field trips, assemblies, teacher inservice days), teachers could only implement the intervention two times in a particular week. In such cases, the teachers 'made up' the session the following week to meet the allotted 15 h of intervention. As mentioned in Section 2.1, intervention groups consisted of 23 to 28 students. Higher-performing readers were paired with lower-performing readers based on screening results and teachers' input about social and behavioral skills. We adopted pairing procedures from other peer-mediated intervention research (Fuchs et al., 1997). Students were rank ordered by reading skill, the list was split in half, the top student from the top half was paired with the top student from the bottom half, and so on. Within each pair, students assumed the roles of "reader" and "helper." Halfway through each session, the roles reversed. Stronger readers read first to provide a model. Pairs changed each week to avoid dependency within pairs.

2.4.2. Texts and questions

Fifty-four narrative texts were developed for the interventions. Some texts were selected from fourth-grade text books and other sources provided by the teachers; other texts were written by project staff (see sample text in Appendix C). Each text had a goal structure with a main goal that spanned the length of the text situation and subgoals related to the main goal. The topic of each text was designed to be age-appropriate and interesting to fourth-graders. The Flesch Reading and Grade Levels were relatively constant across text. Each text was printed, 1.5-spaced, on 8 in. × 11 in. paper in size 14 Times New Roman font. Questions were developed within each text to fit the parameters of the three questioning interventions (sample causal questions are provided in Appendix C).

2.4.2.1. Causal questions. We used the causal networks of the texts to determine where a causal inference might be generated that would help the reader construct a coherent representation of the text. Questions were designed such that their answers would indicate whether the appropriate causal inference was generated. Correct and incorrect answers were identified and included in the intervention protocol to aid teachers and students in providing feedback to readers. On average, six questions were included in each text.

2.4.2.2. General questions. The following question was interspersed at regular intervals throughout the texts (every 5–6 sentences to correspond with the number of questions asked in the Causal intervention): "How does the sentence you just read connect with something that happened before in the story?" Once students were familiar with this question, it was shortened to "Connect it!" to be more efficient.

2.4.2.3. W-questioning. "Who," "what," "where," and "when" questions were interspersed at regular intervals throughout the text, consistent with the number and locations of questions used in the Causal and General conditions. W-questions were designed to have answers that could be located in the last sentence the participant read.

2.4.3. Intervention protocol

In a typical intervention session, students read one or two stories. The teacher read a brief segment of text aloud, while pairs followed along on their own copies. The teacher stopped reading when the first question was to be asked, indicated in the students' text with a large, bolded letter. At this point, the first readers re-read the text aloud to their partners. Next, the teacher asked the question aloud. Readers told the answer to their helpers, and the helpers guided the readers to the correct answer using prompts on answer sheets, which varied according to condition. In the Causal and W-questioning conditions, the helper prompted the reader with, "What else does it say in the story about *[original specific question]*?" If the answer was still not accurate, the helper told the reader the answer. In the General questioning intervention, there were many possible correct answers. Thus, the pair came to an agreement as to whether the answer was appropriate based on the preceding text information.

After all students finished answering the first question, the teacher read the next text segment, and repeated the procedures described above. After 10 min of reading and answering questions, the students switched roles and continued the process for another 10 min. Students earned points for reading and answering questions; points were tallied and reported at the end of each week, and pairs were recognized for their hard work.

2.5. Procedures

2.5.1. Training project staff

Project staff consisted of 12 graduate students in Educational Psychology, one of whom served as a project coordinator, and three undergraduate Psychology majors. All had previous experience working with school-aged children. In September and October of the study, the project coordinator trained all staff to administer the screening and pre-test measures in three separate 2-hour sessions. An additional 2-hour session was held prior to posttesting (in January) to review test administration procedures.

2.5.2. Test administration

Project staff group-administered the screening measures in one session in each class. Classroom teachers were present to help monitor student behavior. Project staff individually administered recall and think-aloud protocols in one session for pretesting, which occurred after screening and before the start of the classroom intervention, and recall protocols for posttesting, which occurred immediately following the classroom intervention.

2.5.3. Coding recall and think-aloud responses

Project staff transcribed students' responses during recall and think-aloud procedures. Then, raters independently parsed recalls into clauses containing noun/verb combinations, and coded 20% of the protocols together. Overall agreement for coding was $K = .89$ for pretest recall and think-aloud and $K = .91$ for posttest recall. Disagreements were resolved through discussion, and then the remaining protocols were divided among the raters to parse and code independently.

2.5.4. Intervention training, implementation, and fidelity

Teachers participated in a half-day training session led by the first author and graduate research assistants. During training, the research team explained the rationale for the study, described the study design, demonstrated how to implement the intervention, practiced with the teachers, and answered questions. Teachers were provided manuals with scripts to train students to implement the intervention; they were asked to follow the content of the scripts but encouraged to use their own words. The teachers trained their students in five 40-min sessions prior to beginning the intervention. The teachers explained and modeled working with a partner, reading the text, asking and answering questions, providing feedback, helping to sound out difficult words, and using the point system.

For each session, students and teachers moved to classrooms designated for their particular intervention group. For example, if a group of students on the same daily schedule shared three teachers, then three questioning conditions would be taught at this time each day (e.g., Group 1 = Causal; Group 2 = General; Group 3 = W-questioning). In this situation, all three teachers would work with all three intervention groups at some point during the intervention (i.e., they would teach a new group one third of the way through the study).

During each session, project staff was present to assist with materials and procedures. Staff also completed a daily checklist of the intervention steps, noting any concerns that arose. Based on this checklist, average intervention fidelity was 82%. Approximately 5 weeks into the intervention, the project coordinator conducted a formal fidelity check in each class, using the same checklist. During these observations, interventions were administered, on average, with 90% accuracy (range: 80%–100%). Staff addressed protocol deviations by coaching students to conduct the steps correctly and prompting teachers to review and reinforce specific procedures.

3. Results

3.1. Effects of questioning interventions for struggling, average, and good readers

We first compared the effects of the Causal, General, and W-questioning interventions on the text recall of struggling, average, and good readers using the five recall variables. Table 2 shows pre- and posttest means and *SD*s by intervention group and reader type.

3.1.1. Pre-treatment comparisons

To determine pre-treatment comparability of the three groups, we conducted a MANOVA using pretest recall as dependent variables. The MANOVA revealed no significant pre-treatment differences between intervention groups, Wilks' Lambda = .944, $F(10, 464) = 1.35$, $p = .202$, and no statistically significant intervention group by reader-type interactions, Wilks' Lambda = .899, $F(20, 770) = 1.26$, $p = .199$. Thus, the three intervention groups appeared to be comparable. As expected, there were significant differences between reader types, Wilks' Lambda = .864, $F(10, 464) = 3.52$, $p < .001$. Follow-up ANOVAs indicated statistically significant differences for unique conservative items recalled ($F[2, 236] = 8.65$, $p < .001$), highly connected items recalled ($F[2, 236] = 4.075$, $p = .018$), and no-match consistent items recalled ($F[2, 236] = 8.786$, $p < .001$). Post-hoc tests using LSD revealed that struggling readers recalled fewer unique conservative items and highly connected items than did average and good readers. Average readers recalled fewer unique conservative items than good readers did. In addition, struggling readers recalled more information that was inconsistent with the text (i.e., no-match inconsistent items) than did both average and good readers.

Table 2
Pre- and posttest means and SDs for proportions of items recalled by questioning intervention and reader type.

Proportions of items recalled to number of story units		Causal ($n=105$)				General ($n=77$)				W-question ($n=63$)			
		Pre-test		Post-test		Pre-test		Post-test		Pre-test		Post-test	
		Mean	(*SD*)	Mean	(*SD*)	Mean	(*SD*)	Mean	(*SD*)	Mean	(*SD*)	Mean	(*SD*)
Conservative items recalled	Struggling[a]	0.281	(0.130)	0.339	(0.144)	0.276	(0.107)	0.406	(0.154)	0.294	(0.141)	0.385	(0.115)
	Average[b]	0.338	(0.104)	0.410	(0.107)	0.361	(0.131)	0.423	(0.122)	0.309	(0.160)	0.383	(0.123)
	Good[c]	0.393	(0.134)	0.489	(0.145)	0.398	(0.149)	0.475	(0.122)	0.381	(0.108)	0.419	(0.099)
Liberal items recalled	Struggling[a]	0.018	(0.017)	0.022	(0.015)	0.019	(0.016)	0.021	(0.024)	0.021	(0.017)	0.027	(0.023)
	Average[b]	0.017	(0.015)	0.024	(0.020)	0.023	(0.021)	0.030	(0.024)	0.018	(0.015)	0.023	(0.020)
	Good[c]	0.020	(0.017)	0.020	(0.016)	0.021	(0.017)	0.029	(0.019)	0.029	(0.030)	0.029	(0.015)
Highly connected items recalled	Struggling[a]	0.067	(0.038)	0.081	(0.035)	0.076	(0.031)	0.087	(0.032)	0.075	(0.038)	0.097	(0.032)
	Average[b]	0.083	(0.031)	0.101	(0.033)	0.088	(0.032)	0.096	(0.036)	0.076	(0.041)	0.094	(0.034)
	Good[c]	0.091	(0.039)	0.109	(0.038)	0.094	(0.038)	0.111	(0.035)	0.091	(0.019)	0.096	(0.032)
No-match, consistent with text	Struggling[a]	0.060	(0.049)	0.057	(0.042)	0.047	(0.035)	0.049	(0.037)	0.068	(0.046)	0.070	(0.043)
	Average[b]	0.061	(0.049)	0.051	(0.034)	0.064	(0.039)	0.060	(0.047)	0.053	(0.035)	0.067	(0.048)
	Good[c]	0.050	(0.034)	0.051	(0.035)	0.069	(0.039)	0.061	(0.037)	0.071	(0.048)	0.052	(0.036)
No-match, inconsistent with text	Struggling[a]	0.039	(0.045)	0.037	(0.036)	0.021	(0.024)	0.031	(0.033)	0.068	(0.056)	0.039	(0.042)
	Average[b]	0.016	(0.016)	0.017	(0.018)	0.027	(0.025)	0.025	(0.024)	0.030	(0.034)	0.019	(0.024)
	Good[c]	0.018	(0.021)	0.010	(0.014)	0.025	(0.017)	0.017	(0.015)	0.020	(0.034)	0.019	(0.021)

[a] Struggling $n=26$ Causal, 16 General, 14 W-questioning.
[b] Average $n=51$ Causal, 40 General, 38 W-questioning.
[c] Good $n=27$ Causal, 21 General, 11 W-questioning.

3.1.2. Intervention effects

To determine intervention effects, we conducted repeated measures MANOVA (RM-MANOVA) with time as the within-subjects factor and intervention group and reader type as between-subject factors. A power analysis (using G*Power 3; Faul, Erdfelder, Lang, & Buchner, 2007) setting power at .80 and anticipating a small effect size ($d=.20$) indicated that our sample size was sufficient to detect statistical significance. The RM-MANOVA revealed a statistically significant main effect of time, Wilks' Lambda = .774, $F(5, 231) = 13.50$, $p<.001$. Follow-up ANOVAs indicated significant effects of time on unique conservative items recalled, $F(1, 235) = 63.46$, $p<.001$; unique liberal items recalled, $F(1, 235) = 4.42$, $p=.04$; no match inconsistent items recalled, $F(1, 235) = 4.16$, $p=.04$, and highly connected items recalled, $F(1, 235) = 13.42$, $p<.001$. On each variable, students' recall improved from pre- to posttest (for the no match inconsistent variable, students included *fewer* inconsistent items). There were, however, no statistically significant interactions between time and intervention group (Wilks' Lambda = .971, $F[10, 462] = .68$, $p=.745$) or time and reader type (Wilks' Lambda = .985, $F[10, 462] = .36$, $p=.963$), nor were there interactions among time, intervention group, and reader type (Wilks' Lambda = .933, $F[20, 767] = .68$, $p=.708$). To summarize, within each questioning condition, all readers made reliable gains over time, but there were no significant differences between intervention groups for struggling, average, and good readers.

3.2. Effects of questioning interventions for subgroups of struggling readers

To investigate our second question—whether the interventions differed in their effectiveness for subgroups of struggling readers—we conducted two additional sets of analyses. First, we conducted cluster analyses on the pre-treatment think-aloud data, following the procedures used in previous research (Rapp et al., 2007; van den Broek et al., 2006), to identify subgroups of struggling readers. Second, after identifying these subgroups, we compared their responses to the three different questioning approaches to determine whether there was an intervention by subgroup interaction.

3.2.1. Cluster analysis based on think-aloud responses

The struggling readers' responses to six think-aloud variables were treated as distinct cases in this analysis. The six measures represented (1) valid inferences (*Elaborative and Predictive*), (2) invalid inferences (*Elaborative and Predictive*), (3) text-based inferences (*Connecting and Reinstatement*), (4) metacognitive responses, (5) evaluations, and (6) paraphrases/text repetitions. We used Ward and Hook's (1963) method, which attempts to minimize the sum of squares of observations within any two clusters formed at each step. Two different clustering solutions were considered (i.e., from a two- to a three-cluster solution). The two-cluster solution was adopted as the best description of the data; the three-cluster solution did not account for additional variability in the data. The two-cluster solution was consistent with our previous research (Rapp et al., 2007; van den Broek et al., 2006), and resulted in two subgroups of struggling readers: elaborators ($n=31$) and paraphrasers ($n=25$). Proportions of elaborators to paraphrasers were similar to proportions found in previous studies (Authors) in which we identified 19 elaborators and 10 paraphrasers in one sample of struggling fourth-grade readers, and 25 elaborators and 14 paraphrasers in another sample. Across studies, the ratio of elaborators to paraphrasers has ranged from 1.25:1.00 to 1.90:1.00.

A MANOVA comparing think-aloud responses of the four types of readers (i.e., the two subgroups of struggling readers, plus average and good readers) indicated a statistically significant main effect of reader type, Wilks' Lambda = .186, $F(3, 239) = 348.69$, $p<.001$. Table 3 provides means, *SD*s, and ANOVA results showing main effects for reader type on the number of valid Elaborative and Predictive inferences, invalid Elaborative and Predictive inferences, and paraphrases and text repetitions.

Follow-up analyses were conducted using LSD, and effect sizes were calculated using Cohen's *d* (difference between means divided by pooled *SD*), to determine how the two subgroups differed from each other and from average and good readers. Findings replicated previous research (Rapp et al., 2007; van den Broek et al., 2006). Specifically, the most distinctive feature of the elaborator group was that they generated reliably more valid elaborative and predictive inferences than did the paraphrasers ($d=1.98$, $p<.001$); there were no statistically significant differences between elaborators and average or good readers on this measure. Elaborators also generated more invalid elaborative and predictive inferences than did average ($d=.67$, $p=.002$) and good readers ($d=.78$, $p=.003$), but not more than paraphrasers did. Further, elaborators generated fewer paraphrases and text repetitions than did paraphrasers, average, or good readers ($d=-1.09$, $-.58$, and $-.68$, respectively; all $ps<.01$).

The most distinctive feature of the paraphraser group was that these readers generated more paraphrases than did elaborators ($d=1.09$, $p=.001$); there were no statistically significant differences between

Table 3
Means, SDs, and F-values comparing two subgroups of struggling readers to average and good readers.

Think-aloud responses	Struggling readers: elaborators ($n=31$)		Struggling readers: paraphrasers ($n=25$)		Average readers ($n=129$)		Good readers ($n=60$)		F (3, 245)	p
	Mean	(SD)	Mean	(SD)	Mean	(SD)	Mean	(SD)		
Valid elaborative and predictive inferences	17.97	(6.12)	7.76	(3.61)	16.33	(8.57)	15.45	(7.11)	10.32	<.001
Invalid elaborative and predictive inferences	1.32	(1.42)	1.00	(2.31)	0.55	(1.08)	0.48	(0.85)	4.18	0.01
Paraphrases and text repetitions	4.61	(4.48)	10.28	(6.02)	8.01	(6.09)	8.58	(6.45)	4.68	0.00

paraphrasers and average or good readers on this measure. Paraphrasers also made fewer valid elaborative and predictive inferences than did elaborators, average, and good readers ($d=-1.98$, -1.07, and -1.22; all $ps<.001$, respectively). Again, these findings replicated our previous research (Rapp et al., 2007; van den Broek et al., 2006).

To determine whether the two subgroups of struggling readers differed only in reading processes (as measured by the think-aloud task) or in reading skill in general (as measured by conventional reading assessments), we compared the two subgroups' performance on the screening measures (Maze and GMRT). A MANOVA revealed no reliable between-group differences on these measures, Wilks' Lambda = .941, $F(2, 52)=1.62$, $p=.208$. Means (*SD*s) on the CBM Maze were 7.67 (3.90) for elaborators and 6.00 (3.57) for paraphrasers. Means (*SD*s) on the GMRT were 17.23 (7.14) for elaborators and 18.20 (9.17) for paraphrasers.

3.2.1. Intervention-by-subgroup differences

We examined whether elaborators and paraphrasers responded differently to the three questioning approaches using the five recall variables. MANOVAS indicated no statistically significant pretreatment differences between intervention groups, Wilks' Lambda = .814, $F(10, 92)=.995$, $p=.454$, or between the two subgroups, Wilks' Lambda = .958, $F(5, 46)=.405$, $p=.843$, and no intervention by subgroup interactions, Wilks' Lambda = .866, $F(10, 92)=.685$, $p=.736$. Thus, groups were comparable at pretest.

We examined whether there were differential intervention effects for elaborators vs. paraphrasers. Table 4 shows pre- and posttest means and *SD*s for the five recall variables by intervention group and struggling reader subgroup. Again, we used RM-MANOVA. A power analysis (using G*Power 3; Faul et al., 2007) setting power at .80 and a moderate effect size ($d=.50$) indicated we had a sufficient sample size to detect significant effects. The RM-MANOVA revealed a statistically significant three-way interaction, Wilks' Lambda = .682, $F(10, 92)=1.94$, $p=.049$. Follow-up RM-ANOVAs indicated time by intervention by subgroup interactions that were statistically significant for unique conservative items recalled, $F(2, 50)=5.45$, $p=.007$, and approached significance for highly-connected items recalled, $F(2, 50)=3.04$, $p=.057$.

We followed up on both of these interactions using two sets of RM-ANOVAs. First, we compared elaborators to paraphrasers within each of the three questioning conditions. Because we expected that elaborators would benefit more from Causal questions, and that paraphrasers would benefit more from General questions, we used one-tailed tests for these two contrasts, and two-tailed tests for W-questioning as we had no a priori expectations for this condition.

In the Causal questioning condition, there were no significant between-group differences on unique conservative items recalled, $F(1, 24)=.918$, $p=.174$ (one-tailed), $d=.35$; however, elaborators made statistically significantly greater gains than paraphrasers did on highly connected items recalled, $F(1, 24)=3.62$, $p=.034$ (one-tailed), $d=.86$. In the General questioning condition, paraphrasers made statistically significantly greater gains than elaborators did on unique conservative items recalled, $F(1, 14)=5.91$, $p=.001$ (one-tailed), $d=1.46$, but not on highly connected items recalled, $F(1, 14)=.981$, $p=.17$ (one-tailed), $d=.40$. In the W-questioning condition, there were no statistically significant differences between elaborators and paraphrasers on unique conservative items recalled $F(1, 12)=.004$, $p=.951$ (two-tailed), $d=.19$ or highly connected items recalled $F(1, 12)=1.48$, $p=.247$ (two-tailed), $d=.49$.

Second, we compared the performance of the two subgroups separately across each condition. For elaborators, there were no statistically significant differences across questioning approaches on unique conservative items recalled, $F(2, 28)=.28$, $p=.758$ or highly connected items recalled, $F(2, 28)=1.31$, $p=.287$. For paraphrasers, there was a statistically significant time by questioning group interaction on unique conservative items recalled, $F(2, 22)=5.91$, $p=.009$. Follow up contrasts using LSD indicated that differences favoring paraphrasers in the General questioning condition over paraphrasers in the Causal questioning condition approached significance, $p=.06$ (one-tailed), $d=1.52$. There was no significant time by condition interaction for paraphrasers on highly connected items recalled, $F(2, 22)=1.83$, $p=.185$.

Table 4
Pre- and posttest means and SDs comparing items recalled by questioning condition and reader subgroup.

Proportions of items recalled to number of story units		Causal ($n=26$)				General ($n=16$)				W-question ($n=14$)			
		Elaborators ($n=14$)		Paraphrasers ($n=12$)		Elaborators ($n=10$)		Paraphrasers ($n=6$)		Elaborators ($n=7$)		Paraphrasers ($n=7$)	
		Mean	(SD)	Mean	(SD)	Mean	(SD)	Mean	(SD)	Mean	(SD)	Mean	(SD)
Conservative	Pre	0.286	(0.130)	0.275	(0.136)	0.285	(0.110)	0.261	(0.112)	0.308	(0.130)	0.280	(0.161)
	Post	0.363	(0.131)	0.312	(0.160)	0.337	(0.148)	0.522	(0.074)	0.396	(0.095)	0.373	(0.140)
Liberal	Pre	0.022	(0.019)	0.013	(0.013)	0.018	(0.014)	0.021	(0.020)	0.021	(0.021)	0.021	(0.014)
	Post	0.027	(0.014)	0.016	(0.013)	0.026	(0.028)	0.013	(0.015)	0.040	(0.020)	0.014	(0.018)
Highly connected	Pre	0.062	(0.041)	0.072	(0.035)	0.080	(0.030)	0.068	(0.036)	0.084	(0.038)	0.065	(0.039)
	Post	0.093	(0.030)	0.066	(0.035)	0.082	(0.027)	0.095	(0.041)	0.088	(0.020)	0.105	(0.040)
No-match, consistent with text	Pre	0.072	(0.056)	0.046	(0.036)	0.056	(0.036)	0.032	(0.033)	0.065	(0.041)	0.070	(0.053)
	Post	0.060	(0.033)	0.054	(0.051)	0.044	(0.041)	0.057	(0.032)	0.047	(0.036)	0.093	(0.039)
No-match, inconsistent with text	Pre	0.042	(0.046)	0.035	(0.044)	0.022	(0.028)	0.021	(0.019)	0.057	(0.041)	0.079	(0.069)
	Post	0.043	(0.036)	0.031	(0.038)	0.029	(0.034)	0.032	(0.035)	0.024	(0.017)	0.053	(0.054)

4. Discussion

In this study, we compared the effects of three questioning approaches (Causal, General, and W-questioning) on the reading comprehension of fourth-graders with different levels of comprehension skills. Specifically, we compared students' performance on their recalls that closely matched story units in the texts that they read (conservative, liberal, and highly connected units recalled) and recalls that went beyond the story units in the text (no match consistent and inconsistent units recalled). As a group, readers made statistically significant improvements in their recall of text under each questioning condition, with no differential effects of the questioning approaches on reading comprehension across struggling, average, and good readers. There was, however, an interaction between type of questioning approach and struggling reader subgroup. Below, we discuss these findings, as well as implications for cognitive theory, reading intervention research, and educational practice.

4.1. Effects of different questioning approaches for different groups of readers

Struggling, average, and good readers' comprehension improved significantly from pre- to posttest, suggesting that different types of questions asked during reading *may* improve readers' recall of text. Readers improved their recall of the text in general (i.e., conservative and liberal items) as well as of events central to the causal network of the text (i.e., highly-connected items). These results suggest that readers improved in their ability to establish the 'right' connections, thereby increasing the coherence of their mental representations of the texts.

The *type* of questioning, however, had no discernable impact on readers' improvement, *as a group*. This finding was somewhat surprising, given that we expected that questions designed to prompt inference generation (Causal and General) would be more effective than literal 'W' questions. One interpretation of this finding is that the type of question does not have an effect on readers' comprehension. However, this interpretation is likely premature because of the possible influence of other factors. For example, it is possible that the effect of differential questioning would occur during reading of more challenging texts than those used in this study, or with text of a different genre. We used narrative texts in this study; findings may be different with informational texts. Also, the effect of different questioning approaches might vary based on question placement. We inserted Causal questions in places where a causal inference was needed, whereas we inserted General and 'W' questions at regular intervals in the text. It is conceivable that question placement confounded our design, and that a comparison of General and W-questions at the same locations as Causal questions would have yielded different results.

Whereas we found no differential effect of questioning type among good, average, and struggling readers *as a group*, we did find differential effects of questioning type for subgroups of struggling readers. Specifically, the Causal questioning approach appeared to be more beneficial for elaborators in improving these readers' recall of highly-connected items, whereas the General questioning approach was more beneficial for paraphrasers in improving their overall recall of events that closely matched those in the text (unique conservative items recalled). There was no differential effect of the W-questioning approach for the two subgroups.

These outcomes are consistent with our expectations for which types of questions would benefit the two subgroups of struggling readers. Elaborators were readers who, during think-aloud tasks, made relatively high proportions of elaborative inferences, but whose inferences were often invalid—that is, not accurate given the content of the text. Causal questions, which were designed to help the reader make connections among important causally-related events in the text, may have helped these readers to focus their attention on relevant connections, as is reflected by their improved performance on recall of highly-connected items (those most central to the causal network of the text). Paraphrasers were readers who, during think-aloud tasks, persistently restated the sentences they had just read, making few or no connections between what they had just read and other parts of the text. General questions appear to have helped these readers to recall text beyond the immediate sentences (cf. McKeown et al., 2009), as is reflected by their improved performance on conservative items recalled; that is, items that closely matched the story units in the original texts.

4.2. Implications for cognitive theory

The observed subgroup-by-intervention interactions have potential implications for cognitive theory. First, findings support the assertion that children with comprehension difficulties form a heterogeneous group (e.g., Cain & Oakhill, 2006; Nation et al., 2002) and do not fit a single cognitive profile. Second, the findings suggest that one source of heterogeneity among struggling comprehenders lies in the way they generate inferences (or fail to do so) during reading. The two subgroups of struggling readers identified in this study—elaborators and paraphrasers—are highly similar to those identified in previous studies (Rapp et al., 2007; van den Broek et al., 2006). The subgroups were formed based on the number and validity of inferences generated during think-aloud tasks. As mentioned earlier, it appears that think-aloud tasks, designed to measure reading *processes*, may reveal differences among struggling readers that are undetected by more traditional measures (Kendeou et al., 2010). Third, the finding that the questioning interventions differentially affected the subgroups suggests that these differences should be considered in the design of specific types of intervention.

4.3. Implications for reading intervention research

As Faggella-Luby and Deshler (2008) and others (e.g., Connor et al., 2004; McKeown et al., 2009) have emphasized, it is likely not appropriate to assume that an intervention that yields promising effects will work for all students. Rather, it is important to identify whether different types of struggling readers respond differently to particular instructional approaches. The current results show that not all struggling readers respond in the same way to instructional interventions. Thus, it appears that identification of subgroups of struggling readers is important for intervention researchers as they seek to answer the question "for whom does it work?"

To continue this work, studies are needed in which subgroups of struggling readers are identified a priori, and assigned randomly to different intervention conditions, to confirm subgroup-by-intervention interactions. It is also important to establish whether the questioning approaches are more effective than a business-as-usual control, whether the timing of questioning matters (e.g., questions asked during vs. after reading), and the extent to which effects of the questioning approaches can transfer to a wide variety of challenging narrative and informational texts that are used on a daily basis in schools. Yet another important question is whether one questioning approach is more beneficial than others, or whether a combination is more useful. It is likely that a variety of inferences are important for comprehending texts, and that the types of inferences needed might vary with different texts. Perhaps a hybrid approach using causal and general questions would better prompt readers to make the 'right' connections.

In addition to examining specific factors related to questioning approaches, instructional conditions related to the intensity of the intervention could also be manipulated in future studies to identify those conditions that are most likely to improve outcomes for struggling comprehenders (Faggella-Luby & Deshler, 2008). Such instructional conditions could include focused manipulations of group size, length and frequency of intervention sessions, duration of the intervention,

motivational components, opportunities to respond, text type and difficulty, and other such variables (cf. Warren et al., 2007).

4.4. Implications for educational practice

Additional work is clearly needed before we can make strong recommendations for practice. At the same time, our results suggest three important considerations for teachers interested in implementing the questioning approaches described in this study. First, our findings suggest that not all students respond to questioning interventions in the same way. In other words, the questioning interventions do not fit a 'one size fits all' approach. Thus, it is critical that teachers monitor students' responsiveness to determine whether or not the intervention is having the intended benefit for individual struggling readers.

Second, the intervention is feasible and efficient: It was implemented in a relatively short time period and classroom teachers were able to execute the activities with fidelity. Teachers reported that they appreciated the time devoted to reading and thinking about text. For example, in a questionnaire designed to elicit teacher feedback, teachers reported that they "felt more students were able to respond to higher level comprehension questions," that there was "less inhibition from low [performing] students to respond," and that "the questions after small portions of reading trains them to think."

Third, reading comprehension requires the coordination of multiple skills and reading processes. The most effective reading comprehension programs are likely to comprise multiple components, such as identifying narrative and expository text structures, activating prior knowledge, self-monitoring, summarizing, and questioning (Faggella-Luby & Deshler, 2008). We view the questioning interventions described as this study as just one part of a comprehensive reading program, and thus recommend that it be integrated with other important comprehension activities, such as those described by the National Reading Panel (NICHD, 2000) and other research syntheses (e.g., Berkeley et al., 2010; Gersten et al., 2001).

5. Conclusion

Although many effective interventions exist, a large number of readers continue to struggle with comprehension. Given the importance of reading comprehension for students' success throughout school and later life, it is critical to understand for whom and under what instructional conditions reading interventions are likely to be most effective (Faggella-Luby & Deshler, 2008). Explicitly connecting cognitive theory to educational intervention research can strengthen knowledge about reading comprehension, including reasons why some readers struggle, and how to best intervene (McKeown et al., 2009; Pressley et al., 2006; Rapp et al., 2007). The findings of this study indicate that such a connection can deepen our understanding of the processes that contribute to success and failure in reading comprehension, and can lead to feasible classroom interventions.

Appendix A. Sample recall text and coded response

Sam and his mother recently moved to a big city. Since Sam was new to the city, he had not met any other children. Sam wanted to make some new friends soon. To pass the time, Sam watched squirrels gather nuts in the local park.

One day, Sam and his mother went to the mall. They met the owner of a store. Sam also met the owner's son. His name was Ben. Both of the boys were friendly, and soon Sam and Ben began to talk about animals. Sam told Ben how he liked to watch squirrels in the park. Ben told Sam that he had seen a baby squirrel sleeping on the ground outside the mall just that morning. Ben said that this sounded weird, because he thought that squirrels usually slept in trees or other more hidden places.

Ben and Sam decided to find out more about squirrels to figure out if the baby was okay. They called the zoo and asked to speak to an animal expert. The expert told them that squirrels do not usually sleep on the ground. Sam and Ben wondered if the squirrel had been abandoned. They decided to call the zoo again and ask the expert for help.

The next day some zoo workers found the squirrel and brought it to the zoo hospital. Sam and Ben were glad they were able to help the squirrel. Sam thanked Ben, and was very happy they'd met at the mall. Sam's move was turning out great.

Sample coded recall

Okay, this kid—a person was moving—moved to a new city (*Conservative: near-literal matches with the original text unit*)

And made a friend (*No match consistent: does not directly match the gist of a text unit but is valid and moderately constrained by the text*)

And then they were playing at the park (*No match inconsistent: does not directly match the gist of a text unit and is invalid or not constrained by the text*)

And he said that I like squirrels (*Liberal: matches the meaning of the original text unit*)

His friend said I have a squirrel by the mall (*Liberal: closely matches the gist of the original text unit*)

And then they called the zoo (*Conservative: near-literal matches with the original text unit*)

To ask why was it laying out on the ground (*No match consistent: does not directly match the gist of a text unit but is valid and moderately constrained by the text*)

And then they went back from the zoo (*No match inconsistent: does not directly match the gist of a text unit and is invalid or unconstrained by the text*)

And helped them the next day. (*Liberal: matches the meaning of the original text unit*)

Appendix B. Think-aloud texts

Practice text: Jimmy and the new bike (adapted from Magliano et al., 1999)

Once there was a boy named Jimmy. One day, Jimmy saw his friend Tom riding a new bike. He wanted to buy a bike. Jimmy spoke to his mother. Jimmy's mother refused to get a bike for him. He was very sad. The next day, Jimmy's mother told him that he should have his own savings. Jimmy wanted to earn some money. He asked for a job at a nearby grocery store. He made deliveries for the grocery store. He earned a lot of money. Jimmy went to the department store. He walked to the second floor. Jimmy bought a new bike.

Test text: Brian's magical skill (adapted from van den Broek et al., 2001)

Brian liked to perform for people. His teacher said that she was going to choose someone to be in a school show that would be in the spring. Brian wanted to be chosen. He sang a song for his teacher. It was not very good. His teacher did not choose him. Brian decided to put on a magic act. He bought some fancy magic cards and a hat. He tried to do some tricks on a table.

The tricks were difficult to do. Brian wanted to learn other kinds of activities. He asked a magician if he would teach him. The magician agreed to teach him. Brian visited every day for a month. He watched how each routine was done and practiced a lot. Brian learned how to perform different kinds of magic. He selected tricks that he could do best. He put together a good act, showing his teacher. He made some pretty flowers come out of her ear and then made the flowers disappear. The magic act was a hit and was selected for the show.

Appendix C. Sample text with questions embedded

Baseball

Alex had just moved to a new town with his family. Alex had left all of his good friends behind. This made him very sad. *A* Alex used to love to play baseball with his old friends. His favorite position was second base. Now he didn't have anyone with whom to play baseball. Oh how he wished he had new friends to play baseball with.

One day, Alex decided to walk to the park. Alex thought he might find someone there who liked to play baseball. When he got to the park, he saw that some children were playing baseball there. Alex decided that he would ask if he could join their baseball game. *B* Alex wanted to talk to the children. So, Alex walked over to them.

The children told Alex that they had just hit their baseball over the fence into Mr. Olson's back yard. The children thought Mr. Olson would be mad about the baseball in his backyard because he was known to be a mean old man. No one wanted to go get the baseball back. Without a baseball, they couldn't continue their game. No one was brave enough to retrieve the baseball. The children had decided they would just all go home. *C*

Alex thought that if he got the baseball back, the children might ask him to play with them. Alex told the children that he would go get the baseball from Mr. Olson. *D* He went to the door of Mr. Olson's house and rang the doorbell. When Mr. Olson came to the door, Alex told him about the baseball in his backyard. Mr. Olson smiled and told Alex to wait while he got the baseball. Mr. Olson came back with the baseball and handed it back to Alex. Alex thanked him and ran back to the baseball field. The children were surprised to hear that Mr. Olson was actually really nice. *E*

Then they asked Alex if he wanted to join the baseball game. Alex was happy. Best of all, Alex got to play second base. He was so excited about playing that position. *F* Alex spent the rest of the day playing baseball with his new friends.

A. *Why was Alex very sad?*

Answer: He had left his good friends behind.

B. *Why did Alex decide to ask if he could join the baseball game?*

Answer: He wanted new friends and someone to play baseball with.

C. *Why did the children decide to just all go home?*

Answer: They couldn't continue their game without a baseball.

D. *Why did Alex tell the children he would get the baseball back?*

Answer: The children might ask him to play with them and he would make new friends.

E. *Why were the children surprised to hear that Mr. Olson was nice?*

Answer: They heard that Mr. Olson was mean.

F. *Why was Alex excited about playing 2nd base?*

Answer: It was his favorite position.

References

Berkeley, S., Scruggs, T. E., & Mastropieri, M. A. (2010). Reading comprehension instruction for students with learning disabilities, 1995–2006: A meta-analysis. *Remedial and Special Education*, *31*, 423–436.

Cain, K., & Oakhill, J. (2006). Profiles of children with specific reading comprehension difficulties. *British Journal of Educational Psychology*, *76*, 683–696.

Chall, J. S. (1996). *Stages of reading development* (2nd ed.). Harcourt Brace College Pubs.

Connor, C. M., Morrison, F. J., & Petrella, J. N. (2004). Effective reading comprehension instruction: Examining child by instruction interactions. *Journal of Educational Psychology*, *96*, 682–698.

Ericsson, K. A., & Simon, H. A. (1993). *Protocol analysis: Verbal reports as data.* Cambridge, MA: MIT Press.

Faggella-Luby, M. N., & Deshler, D. D. (2008). Reading comprehension in adolescents with LD: What we know; what we need to learn. *Learning Disabilities Research & Practice*, *23*, 70–78.

Faul, F., Erdfelder, E., Lang, A. -G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, *39*, 175–191.

Fuchs, L. S., & Fuchs, D. (1992). Identifying a measure for monitoring student reading progress. *School Psychology Review*, *21*(1), 45–58.

Fuchs, D., Fuchs, L. S., Mathes, P. G., & Simmons, D. C. (1997). Peer-assisted learning strategies: Making classrooms more responsive to diversity. *American Educational Research Journal*, *34*, 174–206.

Gersten, R., Fuchs, L. S., Williams, J. P., & Baker, S. (2001). Teaching reading comprehension strategies to students with learning disabilities: A review of research. *Review of Educational Research*, *71*(2), 279–320.

Goldman, S. R., & Varma, S. (1995). CAPping the construction-integration model of discourse comprehension. In C. A. Weaver, S. Mannes, & C. R. Fletcher (Eds.), *Discourse comprehension: Essays in honor of Walter Kintsch* (pp. 337–358). Hillsdale, NJ: Erlbaum.

Goldman, S. R., & Varnhagen, C. K. (1986). Memory for embedded and sequential story structures. *Journal of Memory and Language*, *25*, 401–418.

Graesser, A. C., & Clark, L. F. (1985). *The structures and procedures of implicit knowledge.* Norwood, NJ: Ablex.

Graesser, A. C., McNamara, D. S., Louwerse, M. M., & Cai, Z. (2004). Coh-Metrix: Analysis of text on cohesion and language. *Behavior Research Methods, Instruments & Computers*, *36*, 193–202.

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.

Kendeou, P., Muis, K. R., & Fulton, S. (2010). Reader and text factors on reading comprehension processes. *Journal of Reading Research*, *34*, 365–383.

Kendeou, P., & van den Broek, P. (2005). The effects of readers' misconceptions on comprehension of scientific text. *Journal of Educational Psychology*, *97*(2), 235–245.

Kintsch, W. (1998). *Comprehension: A paradigm for cognition.* New York: Cambridge University Press.

Kintsch, W., & van Dijk, T. A. (1978). Toward a model of text comprehension and production. *Psychological Review*, *85*, 363–394.

Laing, S. P., & Kamhi, A. G. (2002). The use of think-aloud protocols to compare inferencing abilities in average and below-average readers. *Journal of Learning Disabilities*, *35*, 436–447.

Linderholm, T., & van den Broek, P. (2002). The effects of reading purpose and working memory capacity on the processing of expository text. *Journal of Educational Psychology*, *94*, 778–784.

MacGinitie, W. H., MacGinitie, R. K., Maria, K., & Dreyer, L. G. (2002). *Gates–MacGinitie Reading Tests fourth edition: Technical report.* Itasca, IL: Riverside Publishing.

Magliano, J., Trabasso, T., & Graesser, A. C. (1999). Strategic processing during comprehension. *Journal of Educational Psychology*, *91*, 615–629.

McGee, A., & Johnson, H. (2003). The effect of inference training on skilled and less skilled comprehenders. *Educational Psychology*, *23*, 49–59.

McKeown, M. G., Beck, I. L., & Blake, R. G. K. (2009). Rethinking reading comprehension instruction: A comparison of instruction for strategies and content approaches. *Reading Research Quarterly*, *44*, 218–253.

Nation, K., Clarke, P., & Snowling, M. J. (2002). General cognitive ability in children with reading comprehension difficulties. *British Journal of Educational Psychology*, *72*, 549–560.

National Center for Education Statistics (2010). The nation's report card: Reading 2009. Retrieved August 2, 2010 from. http://nces.ed.gov/nationsreportcard/reading/

National Institute of Child Health and Human Development (2000). *Report of the national reading panel.* (NIH Publication No. 00-4754). Washington, DC: U.S. Government Printing Office.

Pressley, M., & Afflerbach, P. (1995). *Verbal protocols of reading: The nature of constructively responsive reading.* Hillsdale NJ: Erlbaum.

Pressley, M., Graham, S., & Harris, K. (2006). The state of educational intervention research as viewed through the lens of literacy intervention. *British Journal of Educational Psychology*, *19*, 1–19.

Pressley, M., McDaniel, M. A., Turnure, J. E., Wood, E., & Ahmad, M. (1987). Generation and precision of elaboration: Effects on intentional and accidental learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *13*, 291–300.

Rapp, D. N., van den Broek, P., McMaster, K. L., Kendeou, P., & Espin, C. A. (2007). Higher-order comprehension processes in struggling readers: A perspective for research and intervention. *Scientific Studies of Reading*, *11*, 289–312.

Snow, C. (2002). *Reading for understanding: Toward an R & D program in reading comprehension.* Santa Monica, CA: RAND Corporation.

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.

Trabasso, T., & van den Broek, P. (1985). Causal thinking and the representation of narrative events. *Journal of Memory and Language*, *24*, 612–630.

van den Broek, P. (2010). Using text in science education: Cognitive processes and knowledge representation. *Science*, *328*, 453–456.

van den Broek, P., Tzeng, Y., Risden, K., Trabasso, T., & Basche, P. (2001). Inferential questioning: Effects on comprehension of narrative texts as a function of grade and timing. *Journal of Educational Psychology*, *93*, 521–529.

van den Broek, P., White, M. J., Kendeou, P., & Carlson, S. (2009). Reading between the lines: Developmental and individual differences in cognitive processes in reading comprehension. In R. K. Wagner, C. Schatschneider, & C. Phythian-Sence (Eds.),

Beyond decoding: The behavioral and biological foundations of reading comprehension (pp. 107–123). New York, NY: The Guildford Press.

van den Broek, P., Young, M., Tzeng, Y., & Linderholm, T. (1998). The landscape model of reading: Inferences and the on-line construction of a memory representation. In H. van Oostendorp, & S. R. Goldman (Eds.), *The construction of mental representations during reading* (pp. 71–98). Mahwah, NJ: Erlbaum.

van den Broek, P., McMaster, K., Rapp, D. N., Kendeou, P., Espin, C., & Deno, S. (2006). *Connecting cognitive science and educational practice to improve reading comprehension.* Washington, D.C: Presented at the Institute of Education Sciences Research Conference.

Ward, J. H., & Hook, M. E. (1963). Application of an hierarchical grouping procedure to a problem of grouping profiles. *Educational and Psychological Measurement*, *23*, 69–81.

Warren, S. F., Fey, M. E., & Yoder, P. J. (2007). Differential treatment intensity research: A missing link to creating optimally effective communication interventions. *Mental Retardation and Developmental Disabilities*, *13*, 70–77.

Wayman, M. M., Wallace, T., Wiley, H. I., Ticha, R., & Espin, C. A. (2007). Literature synthesis on curriculum-based measurement in reading. *Journal of Special Education*, *41*, 85–120.

Williams, J. P. (1993). Comprehension of students with and without learning disabilities: Identification of narrative themes and idiosyncratic text representations. *Journal of Educational Psychology*, *85*, 631–641.

Yuill, N., & Joscelyne, T. (1988). Effect of organizational cues and strategies on good and poor comprehenders' story understanding. *Journal of Educational Psychology*, *80*, 152–158.

Yuill, N., & Oakhill, J. (1988). Effects of inference awareness training on poor reading comprehension. *Applied Cognitive Psychology*, *2*, 33–45.