An Intervention to Improve Comprehension of Cause/Effect Through Expository Text Structure Instruction

Joanna P. Williams, Simonne Pollini, Abigail M. Nubla-Kung, Anne E. Snyder, Amaya Garcia, Jill G. Ordynans, and J. Grant Atkins
Teachers College, Columbia University

This study evaluated the effectiveness of an intervention for second graders at risk for academic failure, which taught reading comprehension embedded in social studies content. The intervention included instruction about the structure of cause/effect expository text, emphasizing clue words, generic questions, graphic organizers, and close analysis of well-structured examples of cause/effect text. It was compared to a program that focused on the same social studies content but without cause/effect training, and to a no-instruction control. Fourteen teachers, randomly assigned to treatment, provided the instruction; 197 7- and 8-year-olds participated. The intervention group demonstrated higher performance than the other groups on both sentence combining and answering comprehension questions. The 2 instructed groups did not differ on the social studies measures, and both were better than the no-instruction group; thus, embedding text structure instruction did not lessen the amount of social studies content acquired. These findings corroborated studies on another text structure (comparison) and extended previous work focused on cause/effect. New findings included, first, more robust group differences in performance than were found in an earlier cause-effect study because of a more precise identification of the instructional level appropriate for this population: the sentence, not the paragraph. Second, examining the sustainability of the intervention effects, a delayed posttest showed that after summer break, the intervention group performed better than the other groups on sentence combining, although not on answering a comprehension question.

Keywords: instruction, reading comprehension, text structure, cause-effect, students at risk for academic failure

The recent emphasis on expository text in primary-level reading comprehension instruction has led to greater exposure to expository text in the classroom, as recommended by the National Research Council (Snow, Burns, & Griffin, 1998). However, there is little to no explicit instruction. This is justified on the grounds that instruction must be designed with an awareness of its developmentally appropriateness, and it accords with the views of many scholars (e.g., Smolkin & Donovan, 2003) who believe that young children have not matured to the point where they can benefit from explicit instruction in comprehension.

But some parts of the primary school curriculum have shown the effectiveness of explicit instruction (e.g., word recognition [National Reading Panel, 2000] and arithmetic [Fuchs & Fuchs, 2003]). Moreover, there is evidence that primary-grade children can benefit from such instruction in the comprehension of narrative text (R. Brown, Pressley, Van Meter, & Schuder, 1996). While it is, of course, essential that instruction be developmentally appropriate, could we not honor that principle and yet also give young children the advantage of explicit instruction in reading comprehension as well as in other subject matters?

Integrating comprehension training and content-area learning provides an opportunity to introduce more expository text into the early elementary curriculum (Guthrie & Ozgungor, 2002; Pappas, Varelas, Patton, Ye, & Ortiz, 2012). Williams (2003) proposed text structure as an instructional focus when integrating reading comprehension and primary-grade content-area instruction. Text structure is key to understanding: Text structures represent rhetorical structures that are fundamental to most if not all comprehension and reasoning (Dickson, 1999). Expository texts, because they typically contain unfamiliar content as well as unfamiliar structure, are often a challenge for children (Goldman & Rakestraw, 2000). Moreover, these difficulties are exacerbated by the fact that textbooks are often dense and poorly structured, so that children run into problems trying to organize and understand the information they are reading (Beck, McKeown, Sinatra, & Loxterman, 1991). A recently developed comprehensive model of the cognitive skills that underlie reading comprehension indicates that the ability to plan and organize, which is based on identifying text structure, is essential to successful middle school reading (Eason, Goldberg, Young, Geist, & Cutting, 2012). There is ample empirical evi-
dence that instruction in text structure improves comprehension at the middle school level and above (Meyer & Ray, 2011; Meyer et al., 2010; RAND Reading Study Group, 2002).

On the basis of these considerations and the results of studies that indicate that primary-grade children are sensitive to expository structure (Danner, 1976; Garner et al., 1986; Lauer, 2002) and, thus, would be amenable to instruction, Williams and her students (Williams et al., 2007) developed and evaluated an intervention that taught second graders the cause/effect structure, embedded in social studies instruction. Significant differences in favor of the students taught the text structure were seen on several comprehension measures, but they were not large effects. Work on an intervention that taught another structure, compare/contrast, embedded in science instruction (Williams, Hall, & Lauer, 2004; Williams et al., 2005; Williams, Stafford, Lauer, Hall, & Pollini, 2009), on the other hand, yielded robust effects. The success of those studies prompted us to attempt another version of the cause/effect intervention.

This Study

The present study reports on the development and evaluation of the later version of the cause/effect program, developed on the basis of the 2007 results, which suggested that the instructional level of the intervention was pitched too high for second-graders. In our redesign of the program we made a major modification: We changed the focus of training to structure at the sentence level instead of at the paragraph level, as will be explained below. We also simplified the instruction about structure by limiting it to one cause/one effect and not including the one cause/multiple effects structure that we had previously used. In addition, we increased the amount of instruction devoted to the basic concept of causality. Other minor changes in wording and format were also made so that the lessons would flow more smoothly.

The intervention is designed to serve as a supplement to regular classroom instruction. Williams et al. (2007) proposed that at the second grade level, appropriate goals for such an intervention would be (a) to ensure that children could use their knowledge of causality in reading as well as they could use it in listening and (b) to lay the foundation for the further development of the knowledge structure of causality via both listening and reading tasks. The instructional model contained elements of both direct and strategic instruction (Kame’enui & Carnine, 1998; Pressley, 1998), a combination which has been shown in meta-analyses to produce the largest effect sizes for at-risk learners (Swanson & Hoshyn, 1998). The model holds that successful instruction is structured, explicit, scaffolded, and intensive. It proceeds systematically from the simple to the complex, and it provides substantial practice at each step. It incorporates meaningful and interesting tasks, including specially designed training materials (texts) that provide simple, clear templates that exhibit instructional points (Williams, 1998; Williams & Pao, 2013).

The core of the instruction consists of first teaching three well-researched strategies: clue words, generic questions, and graphic organizers. An extensive literature in text processing research, reviewed by Goldman and Rakestaw (2000), underscores the importance of these strategies for teaching text structure. Each of them has been shown effective in previous research in helping readers with both identifying global discourse structure and establishing local coherence: clue words (Loman & Mayer, 1983; Lorch, Lorch, & Inman, 1993); generic questions (Carnine & Kinder, 1985; Jenkins, Heliotis, Stein, & Haynes, 1987), and graphic organizers (Armbruster, Anderson, & Ostertag, 1987).

Second, there is one unique component of the intervention: the close analysis of short paragraphs of well-structured text that embody the characteristics of a particular text structure. This analysis provides the student with an opportunity to become familiar with basic structural patterns, which leads to a strengthening of the mental representation of the specific structures. In contrast, the emphasis in many classrooms is on authentic text, which often is not well structured. This core instruction (the strategies and the text analysis) are incorporated into lessons that include instructional components common to many comprehension programs, including trade book reading by the teacher followed by discussion (A. L. Brown, Palinscar, & Armbruster, 2004; Ruddell, 2006), vocabulary (Beck, Mckewon, & Kucan, 2008; Coyne, Simmons, & Kame’enui, 2004), summarizing (Pearson & Fielding, 1991; Taylor, 1982), and writing (Graham & Harris, 2003; Pressley & McCormick, 1995).

Content

The comprehension instruction is embedded in social studies content, which lends itself to opportunities for cause/effect analysis; in fact, causality is central to the understanding of history and social studies (Beck, Mckewon, & Gromoll, 1989; Coffin, 2004). There is firm evidence that middle school and high school students improve their understanding of history and social studies after instruction in the cause-effect structure (e.g., De La Paz, 2005; Englert, Okolo, & Mariage, 2009). Ferretti and De La Paz (2011) suggested that similar instructional approaches would be feasible in the upper elementary grades. However, apart from the Williams et al. (2007) study, there has been no such research at the primary grade level.

Developmental Considerations

Before children learn to read, they are developing the language and discourse patterns that are basic to both oral and written language (Duke & Carlisle, 2011; Lawrence & Snow, 2011). The concept of causality, like other basic concepts, develops initially through perception and action; young children observe co-occurrences and relations between events and discern the reasons for the co-occurrences (Noordman & Vonk, 1998). Infants as young as 7 months demonstrate sensitivity to the causal structure of events (Saxe, Tzelnic, & Carey, 2007); 2-year-olds can make causal predictions and provide causal explanations (Gopnik et al., 2004). Children’s conceptual understanding of causality increases with their experience and developing language ability (Cain & Oakhill, 2007; Geva, 2006).

Thus, when children are beginning to read, they have a rudimentary grasp of causality. The challenge is to tap into their underlying knowledge and map textual information onto their existing representation of causal structure. At a point where they understand oral discourse, they may not yet be able to understand the expression of complex ideas in the syntax of written discourse, because of the differences between the oral and written registers. Even if the written version replicates exactly the oral version,
CAUSE/EFFECT TEXT STRUCTURE INSTRUCTION

children may still have trouble because they are not sufficiently automatic with word-level reading processes, i.e., decoding and general vocabulary. Also, children’s grasp of the concept of causality might not be quite firm, and/or they may not be thoroughly familiar with critical terminology. Moreover, young readers have working memory constraints (Berninger et al., 2010; Swanson & Berninger, 1995). They may be able to comprehend a certain amount of information if it is presented concisely but not the same information when it occurs along with other input. They are also likely to have more difficulty with texts that require inferencing across pieces of text or integrating text and background knowledge (Kintsch & Kintsch, 2005).

Knowledge of the structure of expository genres is typically not complete at the end of middle school (Geva, 2006) or even high school (Goldman & Rakestraw, 2000). Several studies have indicated that cause/effect is a particularly difficult structure (Chevront, 2002; Ciardiello, 2002; Richgels, McGee, Lomax, & Sheard, 1987). While developmental trajectories that trace progress toward mature comprehension of causal structure have not been fully established, there is considerable evidence of children’s gradual improvement in their understanding and in their ability to use the language of causality (Crosson, Leseaux, & Martiniello, 2008; Geva, 2006; Goldman & Murray, 1992). This evidence is helpful in identifying suitable foci for instruction at various points along the developmental continuum.

What would an appropriate focus be for second graders? Most of the text structure research to date has focused on text, i.e., the level of language in which the organization of a series of sentences represents the structure (Meyer & Poon, 2001; Williams et al., 2009). Kintsch’s theory of comprehension (Kintsch, 1998) describes the effects of structure on the identification of important information in a text and on memory: The ordering of words and sentences, as well as specific rhetorical and graphic signals, helps readers select the information that represents the text macrostructure, organize the information into a coherent mental representation, and integrate the textual information with prior knowledge (Mautone & Mayer, 2001).

However, another level of language, the sentence, also can exemplify causal structure (Geva, 2006). A single sentence with multiple clauses can express a complete causal event, both cause and effect. The sentence may also include a connective, to cue coherence. By Grade 2 children have some knowledge of connectives (and other cohesive devices), but they may not be able to use them effectively in text without instruction and practice (Goldman & Murray, 1992). Lack of understanding of connectives at the sentence level has been shown to contribute to difficulties in reading comprehension in 7- to 9-year-olds (Cain, Patson, & Andrews, 2005), older students in Grades 5 and 7 (Geva & Ryan, 1985), and language minority children (Crosson et al., 2008).

Support for organizing instruction at the sentence level comes from the prevalence of sentence-level tasks used in primary-grade research, e.g., Berninger’s work on silent reading comprehension fluency (Berninger, Abbott, & Alsdorf, 1997) and composing (Berninger, Nagy, & Beers, 2011), which found that children from grade one to grade four relied heavily on single sentence compared to multisentence text construction. Also, Crosson et al. (2008) found that the most discriminating assessment items for language minority fourth graders involved sentence-level, and not discourse-level, tasks. Proponents of functional linguistic analysis also emphasize the value of focusing on the single sentence in instruction, which fits with their general recommendation to start with simple language structures in simple contexts and then gradually increase the complexity of the structures (Coffin, 2004; Schleppegrell, Achkugar, & Oteiza, 2004; Unsworth, 2005).

We propose, based on the literature, including our own previous work (Williams et al., 2007), in which we found that second graders performed better on sentence level tasks than on paragraph level tasks, that at that grade level, children at risk for academic failure have reached a point where the sentence is an appropriate level of language for instruction in cause-effect structure. We propose that instruction and practice on individual sentences that at first include connectives (clue words) that function to signal the causal relationship among the constituent propositions will help children move toward the ability to process a sentence that does not include explicit clue words, as well as (although beyond the scope of the present instructional program) a multisentence text, in which the causal structure extends across sentence boundaries.

Purposes of This Study

The main purpose of this study was to modify the Williams et al. (2007) program so that it would more precisely address the level of second-grade students at risk for academic failure and provide a more developmentally appropriate context for determining the value of explicit instruction in text structure for these students, as well as to evaluate the effectiveness of the program compared to a program that taught the same social studies content without the training in text structure and to a no-instruction control.

The second purpose was to investigate the sustainability of the effects of the instruction. An important question in instructional development is whether the effects of an intervention are maintained after its termination. Meta-analyses of studies have shown substantial decreases in achievement scores over the extended summer break (Borman, Benson, & Overman, 2005; Cooper, Nye, Charlton, Lindsay, & Greathouse, 1996). The losses in reading scores are especially pronounced for students in high-poverty schools. Would this text structure intervention be potent enough to maintain its effects? Answering this question is a challenge, given the substantial attrition in grade-school samples over the summer months. In the September following the June completion of the intervention, we administered an abridged version of the posttest to all those students in the sample who were still enrolled in the same schools, in order to determine whether the text structure intervention had sustained effects.

Method

Participants

Participants were drawn from three elementary schools in New York City. The three schools were similar in terms of demographics; all three were categorized as Title 1 schools. The total enrollment across the schools included 76.4% Hispanic, 22% African American, 1.2% Asian or other, and 0.4% Caucasian. Ninety percent of the students received state aid in the form of free or reduced lunch, and approximately 5% of the students were enrolled in either part-time or full-time special education services.
Fifteen teachers volunteered to participate. Their classrooms were randomly assigned to one of three experimental conditions: a text structure program, a content program (focused on the same social studies content but with no cause/effect training), or a no-instruction control, with the restriction that there be twice as many classrooms in each of the instructed conditions as in the no-instruction condition. Conditions were blocked by school. One content group classroom was dropped because the teacher had to stop teaching the program midyear due to scheduling conflicts. Thus there were six text structure classrooms, five content classrooms, and three no-instruction classrooms in the study.

A total of 313 students were enrolled in the 14 participating classrooms. Parental consent forms were returned by 213 students, and of those, 197 students completed the study (the other 16 were reassigned to other classrooms or moved out of the school district during the study). Twenty-one of the 197 students had individual education plans (IEPs) from the onset of the study: nine in the text structure group, six in the content group, and six in the no-instruction group. Additionally two students were referred during the study for IEP evaluation (one in the text structure group and one in the content group). All of the students in the 11 text structure and content classrooms received instruction. In the three no-instruction classrooms, pretests and posttests were administered, but there was no access to any of the instructional materials. Only the students with parental permission were given the pre- and posttests. The text structure condition (TS) included 86 students (mean number per classroom = 14.3, SD = 2.5); the content condition (C) included 63 students (mean number per classroom = 12.6, SD = 1.95), and the no-instruction control (N) included 48 students (mean number per classroom = 16.00, SD = 1.0). There was no significant difference in the mean number per classroom across the experimental conditions, F(2, 11) = 2.53.

Four teachers in the text structure condition, four teachers in the content condition, and two in the no-instruction condition held masters degrees. There were no significant differences among the three treatment conditions in mean number of years of teaching experience: 7.8 (SD = 5.8) in the TS condition, 11.8 (SD = 12.4) in the C condition, and 7.0 (SD = 5.7) in the N condition. Teachers were provided with all of the necessary materials to carry out the lessons and were allowed to keep the materials at the completion of the study. The teachers in the no-instruction condition were given the text structure materials at the conclusion of the study. Teachers were given a small honorarium for their participation.

**Materials**

Our intervention was supplemental to the regular school curriculum. Several different reading programs were used across the schools. There was no social studies program in place; teachers used a range of materials focused on New York City and the local neighborhood. An overview of the major elements in the text structure intervention and the content program is presented in Figure 1. Our focus on text structure and comprehension strategies was in line with the New York State standards for the second-grade language arts curriculum. Our content goal met the New York State standards for the second-grade social studies curriculum. Students learned about three historical communities in the United States: Cherokee, Colonists, and Pioneers. They learned about homes, schools, jobs, and other features in these three communities.

Materials included a trade book and a biography for each of the three communities; selection of the books was based on appropriateness of content and quality of photographs. There also were posters outlining each of the comprehension strategies, a pocket chart to hold picture cards and sentence strips, and posters highlighting the content of the program. Nine well-structured cause-effect paragraphs were written specifically for the intervention. Each target paragraph contained a main idea sentence. There were also several causal sentences, each of which included a cause, an effect, and a clue word. Thus the cause-effect structure was complete within each sentence. The paragraph itself was structured such that all the causal sentences represented specific instances of the topic sentence, providing both structure and content links across the paragraph (Williams, 1984). Noncausal sentences were added to the target paragraphs starting in Lesson 10. The texts required a minimum of inferencing, whether of the gap-filling type necessary to construct a coherent text-base or of the type that requires more background knowledge than would be available from reading the textbooks featured in the lessons (Kintsch & Kintsch, 2005). Examples of the paragraphs used are as follows:

> From Lesson 7:

*Cherokee and Their Jobs*

<table>
<thead>
<tr>
<th>THE TEXT STRUCTURE INTERVENTION</th>
<th>THE CONTENT PROGRAM</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Common Elements</strong></td>
<td></td>
</tr>
<tr>
<td>Trade book reading and discussion</td>
<td>Background knowledge</td>
</tr>
<tr>
<td>Vocabulary development</td>
<td>General content discussion</td>
</tr>
<tr>
<td>Reading target paragraphs</td>
<td>Web graphic organizer</td>
</tr>
<tr>
<td>Comprehension questions</td>
<td>Community notebook</td>
</tr>
<tr>
<td>Lesson Review</td>
<td></td>
</tr>
<tr>
<td><strong>Different Elements</strong></td>
<td></td>
</tr>
<tr>
<td>Clue words</td>
<td></td>
</tr>
<tr>
<td>Cause-effect questions</td>
<td></td>
</tr>
<tr>
<td>Cause-effect graphic organizer</td>
<td></td>
</tr>
<tr>
<td>Analysis of the target paragraphs</td>
<td></td>
</tr>
</tbody>
</table>

*Figure 1.* Overview of the major elements in the text structure intervention and the content program.
There was a lot of work to do in a Cherokee village. Girls were very busy because they had to make baskets and clay pots for storing food. Boys needed to make bows and arrows, since hunters needed these tools to do their job. It was very hard to build a house, so everyone in the village needed to help.

From Lesson 17:

**Pioneers and Their Homes**

The pioneers built their homes from things around them. There were few trees on the prairie, so pioneers made sod houses from grass and dirt. Most homes had dirt floors because wood cost a lot of money. Sometimes the floors were covered with hay, too. Some pioneers didn’t like having dirt walls; therefore, they covered the dirt walls with newspapers. Pioneers sometimes made beds from furry animal skins.

The readability level of the target paragraphs was between third and fourth grade (M = 3.2, SD = 0.44) using the Dale-Chall Readability Scale (Chall & Dale, 1995)—rather high, because some of the words in the text (words relevant to the program content and some of the clue words) are unfamiliar words on the Dale-Chall reading list. However, the students in both the text structure and content programs were quite familiar with these words because they were taught as vocabulary words. When we calculated the readability level with our vocabulary words considered as familiar words, the mean readability of the target paragraphs dropped to between the second- and third-grade level (M = 2.3, SD = 0.5). Five second-grade teachers validated the target paragraphs. They judged individual sentences as main idea, cause-effect, or distractor sentences; 99% of their judgments corresponded to the intended category.

In addition to the target paragraphs, other sentences were written for instruction that conformed to the same pattern as the sentences in the paragraphs (one cause, one effect, and a clue word). Some of these sentences included social studies content, as in the paragraphs. They judged individual sentences as main idea, cause-effect, or distractor sentences; 99% of their judgments corresponded to the intended category.

In addition to the target paragraphs, other sentences were written for instruction that conformed to the same pattern as the sentences in the paragraphs (one cause, one effect, and a clue word). Some of these sentences included social studies content, as in the paragraphs. Familiar content was used in other sentences, e.g., “Jason was very tired so he went to bed early.”

While the focus of the intervention was on reading, we built in a substantial number of writing activities to support and enhance comprehension and learning (Bangert-Drowns, Hurley, & Wilkinson, 2004; Graham & Hebert, 2010). The main writing activities were based on the target paragraphs used in the text analysis and included sentence completion, sentence generation and question answering.

**Text Structure Program**

**Overview.** (Note: In the description of the lesson parts that follows, all the major elements that appear in Figure 1 are represented but not necessarily in the same order. Also, sometimes there are two or more lesson parts that are subsumed under a single major element in the figure.) The instructional program contained three units, each of which focused on one historical community. There were 22 lessons (45 min each) in all; two lessons were taught each week. First, there were two introductory lessons in which the concept of cause/effect was introduced. *Effect* was defined as a thing or event that happens, and *cause* as the thing that makes the event happen. Students developed their understanding of these concepts through role playing activities and picture card activities. In these two introductory lessons, students were also introduced to the program content through two narrative books, *Country Kid, City Kid* (Cummins, 2002) and *On the Town: A Community Adventure* (Caseley, 2002), which follows a boy and his mother as they explore their community.

There were six lessons in Unit 1 (Cherokee) and seven lessons in each of the other two units (Colonists and Pioneers). In the first lesson of each unit (Lessons 3, 9, and 16) the new community was introduced via a biography of a historical figure living during that time period. This lesson included a read-aloud and discussion, an introduction to the relevant vocabulary, and a cause/effect activity. The six additional lessons in both the Colonist and the Pioneer units consisted of two lessons on each of the three features (homes, schools, and jobs). For the first unit, Cherokee, there were five lessons, with only one lesson on the schools feature. Each pair of lessons on a particular feature was organized around a target paragraph and included further read-alouds from a second trade book.

The first lesson on each feature included instruction on the following:

1. **Discussion of cause and effect**
2. **Cause/effect questions.** The students learned two generic questions that could be used to guide them as they analyzed the causes and effects in sentences and in the target paragraphs. These questions were designed to help them focus on the cause/effect information in the text, rather than on irrelevant details or prior knowledge. The two questions were (a) What happened? and (b) Why?
3. **Vocabulary.** The vocabulary concepts were related to the content of the program: community, home, school, job, Native Americans, Cherokee, asi, home school, symbol, hunter, colony, colonist, colonial, keeping room, dame school, blacksmith, pioneer, prairie, sod, sod house, schoolhouse, and farmer. The vocabulary words were embedded in the target paragraphs; they were also presented as a list of words that were explained and illustrated through examples.
4. **Cause/effect clue words.** In Lesson 3, students were introduced to two cause/effect clue words—*because* and *so*. As the program progressed, they were introduced to two additional cause/effect clue words—*therefore* (in Lesson 6) and *since* (in Lesson 7). The students memorized the clue words and used them to create cause/effect sentences and to identify paragraphs as cause/effect paragraphs.
5. **Target paragraph read-aloud and analysis.** Students read the target paragraph silently and then aloud. The class analyzed cause/effect in each sentence by circling the clue word and underlining the parts of the sentence that came before and after the clue word. Students circled each cause clue word (*because* or *since*) using a blue crayon. Students circled each effect clue word (*therefore* or *so*) in green. The part of the sentence that followed the cause clue word (which we call a clause) was identified as the cause, and students underlined this part in blue. The part of the sentence that came before the clue word was identified as the effect, and students underlined this in green. For example, to analyze the sentence “Girls were very busy, because they had to make baskets and clay pots for storing food” students would first circle the clue word “because” in blue. Then they would underline the clause that came after the clue word (“they had to make baskets and clay pots for storing food”) in blue. Finally, the students would underline the clause preceding the clue word (“Girls were very busy”) in green. A comparable...
procedure was followed for sentences containing an effect clue word. Throughout the program, blue was used to indicate cause and green to indicate effect. Sentences that included noncausal information were identified as irrelevant to the causal analysis.

6. Graphic organizer (cause/effect structure). After each cause/effect sentence in the paragraph was analyzed using clue words and underlining, students represented the sentence visually using a graphic organizer. This organizer consisted of two ovals connected by an arrow that indicated the direction of causality (from cause to effect).

Students used their analyzed sentences to complete a graphic organizer for each cause/effect sentence. The cause part of each sentence (underlined in blue) was written in the blue oval on the left and the effect part (underlined in green) was written in the green oval on the right.

7. Lesson review. Each lesson ended with a review of the strategies (clue words, cause/effect questions, and graphic organizer) as well as the content covered.

The second lesson on each feature included instruction on the following:

1. Cause/effect activities using familiar content. The concept of cause/effect was emphasized and developed throughout the program through the inclusion of a variety of activities. These included cloze, matching, sentence combination, picture story, picture puzzle, and cause/effect sentence generation activities.

Cloze activities presented students with incomplete cause/effect sentences and required them to fill in a missing clue word or missing sentence part. Matching activities required students to match causes and effects. Sentence combination activities consisted of using a clue word to combine two sentences into a complete cause/effect sentence. Picture story activities presented students with two boxes, one of which was labeled cause and the other effect. In one of the boxes was a picture and a sentence describing the picture (e.g., a picture of a woman diving into a pool, paired with the sentence “Sara dove into the pool.”). Next to this was a blank box with lines underneath. Students had to generate an appropriate effect for the cause (“Sara dove into the pool”). They drew a picture and wrote a sentence describing the effect that they had generated (e.g., “She made a big splash.”). After this, the students were asked to insert a clue word between the two sentences, thus making a complete cause/effect sentence. Picture puzzle activities were similar; students were presented with two complete pictures and were required to determine which was the cause and which was the effect, then to generate sentences describing each picture, and finally to combine the sentences with a clue word to make a complete cause/effect sentence. Cause/effect sentence generation activities were similar to the picture activities except that they involved no pictures, only sentences.

2. Vocabulary review. Vocabulary words were reviewed. They were presented as a list and were explained and illustrated through examples.

3. Trade book read-aloud and discussion. Specific passages from a trade book were read aloud. Students were encouraged to ask questions and share comments.

4. Community chart. There was a community chart for each unit, which was used for further review of the content. Each chart listed the corresponding home, school, or job vocabulary word that was taught for each specific community. For example, in the Cherokee community chart the words asi, home school, and hunter were written underneath the corresponding feature. During the second and third units, words from previous charts were reviewed.

5. Comprehension questions. Students answered three types of questions about the target paragraph. The first was noncausal (e.g., “What were the different jobs that men and women had to do?”). The second type asked about causes (e.g., “Why was farming a hard job?”), and the third asked about effects (e.g., “What happened because many families did not have very much money?”). We showed the students that they could find the answers to the comprehension questions in the completed graphic organizer as well as in the target paragraph. The students were encouraged to answer in complete sentences.


Content Program

Overview. This program taught the same social studies content and used the same materials (target paragraphs, trade books, and charts) as the text structure program, but it did not focus on cause/effect structure. As in the other program, there were 22 lessons (45 min each; two were taught each week): two introductory lessons; three lessons introducing each historical community via the biography; and two lessons on each of the three features within each unit, with the exception of the Cherokee unit, which included only one lesson on schools.

The two introductory lessons focused on the general content of the program via On the Town: A Community Adventure (Casey, 2002) and City Kid, Country Kid (Cummins, 2002). In these lessons students were exposed to content information about communities and were asked to reflect on the content of the books and to think about different aspects of their own community. In Lesson 3, the Unit 1 biography was presented; the lesson included a read-aloud, a discussion, and a community notebook activity.

The first lesson on each feature included instruction on the following:

1. KWl chart. Comparison lessons began with a discussion of the students’ background knowledge about the community (Cherokee, Colonist, or Pioneer) and the features (home, school, or job) that were the foci of the lesson. This was done through a KWl procedure (what I know, what I Want to know, and what I learned; Ogle, 1986). The teacher recorded responses to “what I know” about a focus feature on a chart. The students then, with the help of the teacher, generated questions about what they wanted to know, and the teacher recorded these. The “what I learned” portion of the chart was completed after the trade book read-aloud, the vocabulary development and the community chart activity. At that point, the students attempted to answer the questions they had generated at the beginning of the lesson and added other details that were discussed during the course of the lesson.

2. Vocabulary. A review of the vocabulary words was conducted as in the text structure program.

3. Trade book read-aloud and discussion. This was the same activity as in the text structure program and also included further discussion questions.

4. Community chart. This was the same activity as in the text structure program.

5. Graphic organizer (content). Students filled in an information web. In the center of the web was a labeled depiction of a
home, school, or job, surrounded by several blank circles, each connected by a line to the center picture. Students filled in the circles with information that they learned during the lesson.

6. Lesson review. Each lesson concluded with a review of the content learned in that lesson, including the features (home, school and job) and the vocabulary.

The second lesson on each feature included instruction on the following:

1. **KWL chart review.** Students reviewed the KWL chart from the previous lesson and were asked if there were additional items they wanted the teacher to add to the “want to Know” or “what I Learned” columns.

2. **Vocabulary review.** A review of the vocabulary words was conducted as in the text structure program.

3. **Read-aloud of the target paragraph.** Students read, first silently and then aloud, the same target cause/effect paragraph that was presented in the text structure program; however, they did not examine the text structure and identify it as a cause/effect paragraph. Rather, the paragraph was presented as another opportunity to read and learn about the community and the feature that was targeted in the lesson.

4. **Comprehension questions.** The students in the content program answered the same noncausal questions that the students in the text structure program answered. However, there were no causal questions; instead, students answered additional questions pertaining to the information in the target paragraph.

5. **Community notebook.** Students drew pictures of something related to the lesson’s topic (e.g., pioneer jobs) and wrote a paragraph about their drawing.

6. **Lesson review.** Each lesson concluded with a review of the content learned in that lesson, including the features (home, school and job) and the vocabulary.

**Procedure**

**Introducing the program to teachers.** All of the second-grade teachers in each school were given a short introduction to the study. We told them that we were trying out two versions of a social studies program, both of which contained a variety of reading and writing activities. We explained that volunteering to participate meant agreeing to be randomly assigned to either one of the two programs or to a no-treatment control. We also asked them not to discuss the study with anyone until after the posttesting and debriefing.

Individual training sessions (30–45 min) were held to familiarize teachers with the program that they were to teach. We discussed the program’s overall goals and reviewed each section of the lessons. These sessions took place about 1 week prior to the beginning of the instruction. Because of the great amount of detail provided in the lessons, there was no need for further teacher preparation. The teachers were asked to tailor the instruction according to their own individual teaching styles and professional judgment but to be sure to deliver all the content and concepts addressed in the lessons as presented in the teachers’ manual.

**Pretest.** The pretest, conducted in two (30–45 min) sessions, consisted of audio-taped individual interviews. In the first session, students were given the Word Identification and Passage Comprehension subtests of the Woodcock Reading Mastery Test, Form H (Woodcock, 1987). The split-half reliability coefficient is 0.97 for the Word Identification subtests, 0.92 for the Passage Comprehension subtest, and 0.97 for the combination of the two subtests.

In the second session, students were given a test to assess their ability to perform several of the tasks to be taught in the instructional program. This second session was audio-taped. One of the four pretest measures (all of which were repeated on the more extensive posttest) assessed the generation of cause/effect questions, a strategy taught in the text structure program. Another measure assessed content knowledge via vocabulary concepts. A third assessed the ability to combine two orally presented sentences into one cause/effect sentence. The fourth measure assessed the ability to answer questions that focused on information acquired from reading a short paragraph.

**Lesson attendance and feedback.** Teachers were asked to take attendance and to complete feedback forms at the end of each lesson.

**Classroom observations.** Observations were conducted in the 11 instructed classrooms to assess fidelity to treatment. Each of the 11 classrooms was observed twice over the course of the instruction, with the exception of one class that was observed only once, due to schedule conflicts. Observers recorded whether each lesson section was completed. Two observers made independent observations of each lesson. Intrarater reliability was 100%. Thereafter, each observation was done by a single observer.

**Posttest.** The posttest consisted of two individual, audio-taped interviews (30–45 min each) and one group session (30 min). During the second individual session, the students were administered Form G of the Word Identification and Passage Comprehension subtests of the Woodcock Reading Mastery Test (Woodcock, 1987).

The three sessions included an extensive array of strategy and outcome measures (see Appendix). The strategy measures included Recalling Clue Words (CLUEWORDS), Recalling the Cause/Effect Questions (QUESTIONS), Completing a Graphic Organizer (GRAPHORG) after reading a short paragraph, and Underlining Clauses (CLAUSEs) in a cause/effect sentence. There was also a fifth measure that assessed a graphic organizer that was given to the content group (WEB).

The **content measures** assessed Feature Questions (FEATURE; information about the features of the three historical communities as presented in the target paragraphs), Additional Information (ADDITION; knowledge of other content presented in the target paragraphs), and Vocabulary Definitions (VOCAB).

The **sentence combination measures** assessed the ability to construct cause/effect sentences when presented with two pictures (SENTPIC), two sentences read by the tester (SENTLISTEN), and two sentences read independently (SENTINDEP).

The **comprehension question measures** required students to answer three types of questions (cause question, effect question, and noncausal question) about a series of paragraphs that for the most part involved social studies content. In the main analyses, performance was based on the cause question and the effect question. The first measure (EXPLICIT) assessed the effects of explicit teaching. This measure was based on two paragraphs that had appeared during the instruction. One of these paragraphs required an oral response and the other, a written response. Next, two novel paragraphs, requiring oral responses, assessed transfer (TRFRORAL). One of these paragraphs contained information about an untaught feature (games) of one of the instructed communities.
and the other information unrelated to the program. There were also two novel paragraphs requiring written responses that assessed transfer (TRFR-WRIT); one contained information about an untaught feature (celebrations) of one of the instructed communities and the other information unrelated to the program. Finally, a paragraph that was taken from a trade book was used to assess transfer to authentic text (TRFR-AUTH).

**Delayed posttest.** The delayed posttest consisted of a shortened version of the original posttest. It included three of the four strategy measures from the original posttest: CLUEWORDS, QUESTIONS, and CLAUSES. All three of the content measures were administered: FEATURE, ADDINFO, and VOCAB. We also added another content measure, Recognition of the Topic (RECOGTOP). Students were given two multiple-choice items to determine whether they remembered the main social studies content presented in the instruction (i.e., the communities and the features).

There were three sentence combination measures, the same as those used earlier: SENTPIC, SENTLISTEN, and SENTINDEP. The comprehension question measure consisted of one novel paragraph that required an oral response (TRFR-ORAL).

**Scoring.** Scoring guidelines were developed for each measure on the basis of 15 randomly selected protocols, five from each of the three experimental treatments. Criteria for judging whether a response was correct, and how many points to give each response, were determined. Two scorers completed independent blind scoring of each measure for 20% of the protocols. Interrater reliability agreement (number of agreements divided by the number of items) ranged from 93% to 100% across measures.

**Teacher debriefing.** After the posttest was completed, individual meetings were held with the 11 teachers who delivered the instruction, to explain the purpose of the study more fully and to receive feedback. The teachers were asked about the benefits of the program and how they would change the program if they were to use it again. Teachers were also given two short written questionnaires: one that asked about their education and teaching experience and one that asked them to rate particular aspects of the program on a scale from 1 to 5. The second questionnaire varied slightly depending on which program (i.e., text structure or content) the teacher had taught. The teachers assigned to the no-instruction group were also debriefed, at which time the purpose of the study was explained, the teacher information questionnaire was administered, and the books and lesson materials used in the instructional program were distributed.

## Results

### Overview

First, information concerning the characteristics of the participants (demographics, reading scores, and attendance) are presented, followed by information about the fidelity of the classroom observations. Then come the results of the evaluation of the program, first, the findings on the immediate posttest and then on the delayed posttest.

### Characteristics of Participants

There were six Text Structure classrooms, five Content classrooms, and three No-Instruction classrooms. The mean number of students tested per classroom was 14.14 ($SD = 2.38$). Statistical analyses were conducted with the classroom as the unit of analysis.

Table 1 presents the characteristics of the participants. No significant differences were found among the three treatment groups (text structure, content, and no-instruction) on age, $F(2, 11) = 0.15$; standard score on the two subsets of the Woodcock Reading Mastery Test: Word Identification subtest, $F(2, 11) = 0.53$; Passage Comprehension subtest, $F(2, 11) = 0.63$, and Total Reading score (the combination of the two subs tests), $F(2, 11) = 0.61$. Mean scores for all pretest measures of the intervention tasks were close to zero. There were no significant differences as a function of instructional condition on any of these measures, and they were not used in our analyses.

The proportion of lessons (out of 22) that each participant attended and the mean proportion for each of the 11 classrooms that received instruction (text structure and content) were cal-

<table>
<thead>
<tr>
<th>Measure</th>
<th>TS</th>
<th>SD</th>
<th>C</th>
<th>SD</th>
<th>N</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gender</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>$N = 38$</td>
<td>7.70</td>
<td>0.22</td>
<td>$N = 26$</td>
<td>7.74</td>
</tr>
<tr>
<td>Female</td>
<td>$N = 48$</td>
<td>7.84</td>
<td>0.23</td>
<td>$N = 37$</td>
<td>7.78</td>
</tr>
<tr>
<td>Age (years)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Word Identification</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Standard score</td>
<td>458.92</td>
<td>15.63</td>
<td>452.70</td>
<td>6.04</td>
<td>460.51</td>
</tr>
<tr>
<td>Grade equivalent</td>
<td>2.6</td>
<td>2.5</td>
<td>2.5</td>
<td>2.5</td>
<td></td>
</tr>
<tr>
<td>Passage Comprehension</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Standard score</td>
<td>475.54</td>
<td>8.32</td>
<td>471.31</td>
<td>4.04</td>
<td>474.90</td>
</tr>
<tr>
<td>Grade equivalent</td>
<td>2.5</td>
<td>2.2</td>
<td>2.2</td>
<td>2.4</td>
<td></td>
</tr>
<tr>
<td>Total Reading Score</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Standard score</td>
<td>467.23</td>
<td>11.57</td>
<td>462.01</td>
<td>4.84</td>
<td>467.70</td>
</tr>
<tr>
<td>Grade Equivalent</td>
<td>2.5</td>
<td>2.3</td>
<td>2.3</td>
<td>2.5</td>
<td></td>
</tr>
</tbody>
</table>

*Note.* TS = text structure; C = content; N = no instruction; WRMT-R = Woodcock Reading Mastery Test—Revised (Woodcock, 1987).
culated. No individual student attended fewer than 59% of the lessons. The mean proportion of lessons attended was .91 ($SD = .03$) for the text structure group and .93 ($SD = .03$) for the content group, with no significant difference between the two groups: $t(9) = 0.99$. Attendance was not recorded for the no-instruction group.

**Fidelity to Treatment**

Across the 21 classroom observations, most of the lesson parts in the text structure program and most of the lesson parts in the content program were taught, not counting the lesson review, which was observed during only one observation. No specific lesson part except for the review was dropped any more frequently than any other part. There were no additions to the lessons of content or strategies beyond what was specified in the program. Since the teachers had been told to use the lessons as a general guideline, rather than as a script (Williams et al., 2005), we considered the fidelity to the lesson outline satisfactory. That is, they followed the lesson plans in the way that they had been directed to do.

**Immediate Posttest**

**Clusters of measures.** Results are presented first for each of the four clusters as a whole. The criterion level for statistical significance was set at $p < .0125$ because there were multiple tests (Bonferroni correction); all four clusters reached significance. Table 2 presents the means and standard deviations of the proportion correct on each of the clusters, as well as $F$s, specific comparisons, and effect sizes. Because the classroom was the unit of analysis in this study, the reported effect sizes cannot be underestimated as they typically are understood, e.g., in terms of Cohen’s (1988) recommendations. Effect sizes based on cluster-level designs and analyses are typically larger than, and not comparable to, individual level effect sizes. It should be noted that because cluster analyses are often underpowered, they tend to yield conservative estimates of statistical significance (What Works Clearinghouse Procedures and Standards Handbook, 2008).

**Strategy cluster.** The strategy cluster was composed of the four measures that evaluated strategies taught to the text structure students. There was a significant difference in performance among the three treatments. As expected, specific comparisons indicated that the text structure group performed significantly better than the content group and the no-instruction group. There was no difference between the content and the no-instruction groups. (The WEB graphic organizer measure was not part of the cluster.)

**Content cluster.** This cluster was composed of the three content measures. Here, the predicted pattern was that the text structure and the content groups would be comparable and that both of these groups would outperform the no-instruction group. The

**Table 2**

<table>
<thead>
<tr>
<th>Measure</th>
<th>TS $n = 6$</th>
<th>C $n = 5$</th>
<th>N $n = 3$</th>
<th>F(2, 11)</th>
<th>Specific comparisons</th>
<th>ES</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Cluster measures</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Strategy</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Recalling Clue Words—CLUEWORDS</td>
<td>.90 .06</td>
<td>.50 .16</td>
<td>.02 .03</td>
<td>.02 .02</td>
<td>***.33.62</td>
<td></td>
</tr>
<tr>
<td>Recalling C/E Questions—QUESTIONS</td>
<td>.58 .33</td>
<td>.50 .14</td>
<td>.03 .02</td>
<td>.03 .06</td>
<td>***.10.66</td>
<td></td>
</tr>
<tr>
<td>C/E Graphic Organizer—GRAPHRG</td>
<td>.18 .13</td>
<td>.01 .03</td>
<td>.01 .01</td>
<td>.01 .01</td>
<td>.50.60</td>
<td></td>
</tr>
<tr>
<td>Underlining Clauses—CLAUSES</td>
<td>.43 .17</td>
<td>.05 .06</td>
<td>.15 .17</td>
<td>.10 .10</td>
<td>***.10.27</td>
<td></td>
</tr>
<tr>
<td>Web Graphic Organizer—WEB</td>
<td>.82 .09</td>
<td>.90 .08</td>
<td>.71 .09</td>
<td>.46.47</td>
<td>**.C &gt; N</td>
<td></td>
</tr>
<tr>
<td><strong>Individual measures</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Content</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Feature Questions—FEATURE</td>
<td>.43 .15</td>
<td>.50 .11</td>
<td>.12 .04</td>
<td>.95.93</td>
<td>***.TS &gt; C, ***TS &gt; N</td>
<td></td>
</tr>
<tr>
<td>Additional Information—ADDMINO</td>
<td>.50 .19</td>
<td>.44 .09</td>
<td>.11 .04</td>
<td>.82.21</td>
<td>***.TS &gt; C, ***TS &gt; N</td>
<td></td>
</tr>
<tr>
<td>Vocabulary—VOCAB</td>
<td>.48 .20</td>
<td>.43 .04</td>
<td>.10 .03</td>
<td>.81.11</td>
<td>***.TS &gt; C, ***TS &gt; N</td>
<td></td>
</tr>
<tr>
<td>Sentence Combination:</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pictures—SENTPIC</td>
<td>.82 .13</td>
<td>.23 .15</td>
<td>.20 .23</td>
<td>.24.48</td>
<td>***.TS &gt; C, ***TS &gt; N</td>
<td></td>
</tr>
<tr>
<td>Read by Tester—SENTLISTEN</td>
<td>.58 .13</td>
<td>.14 .10</td>
<td>.15 .13</td>
<td>.22.70</td>
<td>***.TS &gt; C, ***TS &gt; N</td>
<td></td>
</tr>
<tr>
<td>Read Independently—SENSENTINDEP</td>
<td>.57 .15</td>
<td>.08 .05</td>
<td>.13 .09</td>
<td>.30.69</td>
<td>***.TS &gt; C, ***TS &gt; N</td>
<td></td>
</tr>
<tr>
<td><strong>Comprehension Questions:</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Explicit Teaching—EXPLICIT</td>
<td>.55 .15</td>
<td>.38 .03</td>
<td>.34 .12</td>
<td>4.45</td>
<td>**.TS &gt; C, TS &gt; N</td>
<td></td>
</tr>
<tr>
<td>Oral Transfer—TRFR-ORAL</td>
<td>.55 .18</td>
<td>.28 .15</td>
<td>.28 .08</td>
<td>5.44</td>
<td>**.TS &gt; C, TS &gt; N</td>
<td></td>
</tr>
<tr>
<td>Written Transfer—TRFR-WRIT</td>
<td>.39 .11</td>
<td>.21 .09</td>
<td>.24 .05</td>
<td>7.33</td>
<td>**.TS &gt; C, TS &gt; N</td>
<td></td>
</tr>
<tr>
<td>Authentic Transfer—TRFR-AUTH</td>
<td>.40 .09</td>
<td>.34 .18</td>
<td>.25 .07</td>
<td>1.35</td>
<td>—</td>
<td></td>
</tr>
</tbody>
</table>

**Note.** TS = text structure; C = content; N = no instruction; ES = effect size; C/E = Cause/Effect. Dashes indicate data were not computed because $F$ was not significant. Effect sizes were calculated according to the formula used by the National Reading Panel (2000):

\[
ES = \frac{M_1 - M_2}{\sqrt{\frac{(SD_1^2 + SD_2^2)}}}
\]

* $p < .05$.  ** $p < .01$. *** $p < .001$.  

...
results were as predicted. There was a significant difference in performance among the three treatment groups. Specific comparisons indicated that the text structure group performed significantly better than the no-instruction group and that the content group performed significantly better than the no-instruction group. There was no difference between the text structure and the content groups.

Sentence combination cluster. This cluster was composed of the three sentence combination measures. There was a significant difference in performance among the three treatment groups. Specific comparisons indicated that, as predicted, the text structure group scored significantly higher than the content group and the no-instruction group. There was no difference between the content and the no-instruction groups.

Comprehension question cluster. This cluster was composed of the four comprehension question measures. The score for each measure was the mean of the scores on two questions, a causal question and an effect question. There was a significant effect of treatment. As predicted, the text structure group scored higher than the content group and higher than the no-instruction group. There was no difference between the content group and the no-instruction group.

Individual measures. The mean and standard deviation of the proportion correct on the individual measures within each cluster, Fs, specific comparisons, and effect sizes are also presented in Table 2.

Strategy measures. Each of the four measures when examined separately showed the same pattern as did the strategy cluster as a whole. The overall treatment effect was significant for CLUEWORDS, QUESTIONS, GRAPHORG, and CLAUSES. Specific comparisons showed that in every case, as predicted, the text structure group scored significantly higher than either of the other two groups, and the content and the no-instruction groups did not differ from each other.

The final strategy measure, WEB (not included in the strategy cluster because it evaluated a strategy taught to the content group), showed an effect of treatment. Specific comparisons indicated that the content group performed at a higher level than the no-instruction group. There was no difference between the content group and the no-instruction groups.

Content measures. Each of the three measures, FEATURE, ADDINFO, and VOCAB, showed the same pattern that the content cluster as a whole showed. In this case, the predicted pattern was that both of these groups would outperform the no-instruction group. The overall treatment effect was significant; specific comparisons indicated that in every case, the text structure and content groups did not differ, and both of these groups scored significantly higher than the no-instruction group.

Sentence combination measures. Each of the three measures in this cluster, SENTPIC, SENTLISTEN, and SENTINDEP, when examined separately, showed the same pattern as the cluster showed. There was an overall significant treatment effect, and specific comparisons indicated that in every case, the text structure group scored significantly higher than either of the other two groups, and the content and the no-instruction groups did not differ from each other.

Comprehension question measures (Table 2). Three of the four measures in this cluster, EXPLICIT, TRFR-ORAL, and TRFR-WRIT, showed a significant overall effect of treatment. In all three cases, specific comparisons showed the predicted pattern: The text structure group scored significantly higher than either the content or the no-instruction group. On the fourth measure, TRFR-AUTH, however, while the treatment means followed the same pattern as the other three, the differences did not reach significance.

Further analysis of the comprehension question measure. For each of the sentences in each of the test paragraphs, the three questions asked for (a) the cause that had been stated in the sentence (Why?), (b) the effect that had been stated in the sentence (What happened?), and (c) noncausal information that tapped knowledge of a paragraph detail that was not associated with the causal information. (See Appendix for examples of these questions.)

For the previous analyses, we chose the mean of the cause questions and the effect questions as the overall measure. Now we looked at each of the three questions separately. Table 3 presents the mean and standard deviation of the proportion correct, as well as Fs, specific comparisons, and ESs for each type of question, for each treatment condition. There was a significant overall effect on the cause question as well as on the effect question. Specific comparisons indicated that, as expected, the text structure group performed significantly higher than the comparison and the no-instruction groups, both on the cause question and on the effect question. There was no effect of treatment on the noncausal question. This pattern of results was as predicted.

Delayed Posttest

Of the original sample of 197 students, 123 were available for the delayed posttest. The mean number of students tested per classroom was 8.71 ($SD = 3.86$). The proportion of students who completed the delayed posttest did not differ among the groups: .63 ($SD = .35$) in the TS group, .55 ($SD = .39$) in the C group, and .67 ($SD = .30$) in the N group). There was no significant difference among the groups, $F(2, 11) = 0.89$. No significant differences

Table 3

<table>
<thead>
<tr>
<th>Question type</th>
<th>TS $n = 6$</th>
<th>C $n = 5$</th>
<th>N $n = 3$</th>
<th>$F(2, 11)$</th>
<th>Specific comparisons</th>
<th>ES</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cause</td>
<td>.50</td>
<td>.34</td>
<td>.33</td>
<td>5.85</td>
<td>$^<em>$TS &gt; C, $^</em>$TS &gt; N</td>
<td>1.68, 2.62</td>
</tr>
<tr>
<td>Effect</td>
<td>.47</td>
<td>.29</td>
<td>.23</td>
<td>7.71</td>
<td>$^<em>$TS &gt; C, $^</em>$TS &gt; N</td>
<td>1.71, 2.09</td>
</tr>
<tr>
<td>Noncausal</td>
<td>.76</td>
<td>.73</td>
<td>.67</td>
<td>1.11</td>
<td>—</td>
<td>—</td>
</tr>
</tbody>
</table>

Note. TS = text structure; C = content; N = no instruction; ES = effect size. Dashes indicate data were not computed because $F$ was not significant. $^* p < .05$. $^*^* p < .01$. 

<table>
<thead>
<tr>
<th>M</th>
<th>SD</th>
<th>M</th>
<th>SD</th>
<th>M</th>
<th>SD</th>
<th>$F(2, 11)$</th>
<th>Specific comparisons</th>
<th>ES</th>
</tr>
</thead>
<tbody>
<tr>
<td>.50</td>
<td>.11</td>
<td>.34</td>
<td>.08</td>
<td>.33</td>
<td>.02</td>
<td>5.85</td>
<td>$^<em>$TS &gt; C, $^</em>$TS &gt; N</td>
<td>1.68, 2.62</td>
</tr>
<tr>
<td>.47</td>
<td>.12</td>
<td>.29</td>
<td>.09</td>
<td>.23</td>
<td>.11</td>
<td>7.71</td>
<td>$^<em>$TS &gt; C, $^</em>$TS &gt; N</td>
<td>1.71, 2.09</td>
</tr>
<tr>
<td>.76</td>
<td>.13</td>
<td>.73</td>
<td>.04</td>
<td>.67</td>
<td>.06</td>
<td>1.11</td>
<td>—</td>
<td>—</td>
</tr>
</tbody>
</table>
were found among the three treatment groups on age, $F(2, 11) = 0.11$; standard score on the two subtests of the Woodcock Reading Mastery Tests, Word Identification, $F(2, 11) = 0.67$; Passage Comprehension, $F(2, 11) = 0.74$; and Total Reading score, $F(2, 11) = 0.28$.

Table 4 presents the means and standard deviations of the proportion correct on each of the clusters and each of the individual measures, as well as $Fs$, specific comparisons, and ESs. The difference in performance across the treatment groups on the Strategy and Content clusters did not reach a conventional level of significance; however, because both $ps$ were under .10 and the pattern of results had been predicted, we did specific comparisons. For the Strategy cluster, the two comparisons that had been predicted to show differences showed trends in the expected direction: TS > C, $p = .058$; and TS > N, $p = .064$. The third comparison, C versus N, was not significant.

For the Content cluster, there was a significant difference between the TS and the N groups, as predicted, but not between C and N. There was no significant difference between the TS and C groups, as expected. On the Sentence Combination cluster, there was an overall significant difference, and the two predicted specific comparisons were also significant, TS > C, and TS > N; there was no difference between C and N, as expected.

**Individual Measures**

Within the Strategy cluster, two of the three measures showed a significant effect of treatment group: CLUEWORDS and QUESTIONS. The two predicted specific comparisons, TS > C and TS > N, were significant for all three of these measures; C was not different from N. On the Comprehension Question measure, while the mean scores of the treatment groups followed the predicted pattern, the differences among groups was not significant.

**Discussion**

This study, which evaluated an intervention that taught cause/effect structure embedded in social studies instruction, showed that (a) explicit comprehension instruction can indeed be effective at the primary level (students who received the text structure intervention outperformed students who received the comparison program and the no-treatment program), and (b) such instruction can be accomplished within the framework of content-area instruction without a loss in the amount of content acquired (there were no significant differences between performance on the two programs, and both were superior to the no-treatment program). While the earlier study (Williams et al., 2007) provided initial evidence of the effectiveness of the intervention, the present findings were
more robust. Here, we addressed more precisely the level of at-risk second graders, and the findings showed clearly that our earlier, more tentative conclusions about teaching causal structure were sound.

The same measures were used in the two studies, with certain exceptions (and the pattern of findings was similar). First, one of the 2007 strategy measures required students to locate clue words in texts; differences between the groups were huge. In the present study, in order to provide a more challenging task, we used a recall task, which yielded a similar outcome. Second, there were no sentence-combination measures in the earlier study. Third, the 2007 comprehension question measures included some one cause/multiple effects questions. Because the new version of the intervention did not teach that structure, those measures were omitted. In the present study the intervention group was significantly superior to the other groups on both the cause and the effect questions, whereas in the earlier study the pattern of results was the same but only the effect questions showed differences that were significant.

These findings, coupled with the results of evaluations of Williams’ intervention to teach the comparison structure within the context of science instruction (Williams et al., 2005, 2009), which also showed robust effects, suggest that a focus on text structure is valuable for young children across a variety of content areas. It also suggests that explicit instruction in general has perhaps not been sufficiently exploited in the primary grades curriculum.

The main modifications made in the program for this study were the focus on cause/effect sentences instead of cause/effect paragraphs, the restriction of the cause/effect structural pattern to one cause/one effect, and the increased emphasis on the basic concept of causality. We believe that it was these changes in the program, consistent with our instructional model, that were responsible for the greater effectiveness of the instruction. This is a more plausible conclusion than the possibility that other, much smaller changes that were made in the lessons were responsible (e.g., the fact that clue words were introduced gradually instead of all at once or that the graphic organizer and the generic questions were simplified in order to make the instruction flow more smoothly). Of course, this study was designed to evaluate the effectiveness of an intact instructional program and does not provide evidence of the effectiveness of any single component of the instruction. It would be useful, for this intervention as well as for others, to explore the issue of componential analysis, to determine which components to emphasize in future iterations of the program. For example, our intervention incorporates a substantial number of writing activities because of the extensive literature that demonstrates the many ways in which writing enhances reading (Graham & Hebert, 2010); it is likely, although this hypothesis cannot be evaluated in the present design, that the writing requirements, which were closely tied to the text analysis, represented a potent factor in the success of the intervention.

The present study extends previous findings by showing that some of the effects of text structure training (on the strategy and the sentence combination measures) remained, to a limited extent, after an extended summer break. In contrast, the social studies content measures did not show any carry-over of instructional effects to the delayed posttest (although the instructed students were able to recognize which communities they had studied). Content measures are the most common type of measure included in studies of the effects of extended periods of no instruction/no practice; thus, this latter finding is not surprising in light of the literature on the topic, which is generally pessimistic about the sustainability of instructional efforts (Borman et al., 2005).

The fact that the sentence combination measures did show a carryover effect was an encouraging finding. It corroborated the conclusions from the immediate posttest and indicated that our instruction achieved some success in moving students forward in their ability to deal with cause/effect structure. This improved ability to grasp sentence-level cause/effect should provide the foundation for the more challenging task of comprehending cause/effect more generally. We did not see significant carryover effects on the comprehension question measure, although that measure did show the predicted pattern of findings. Given the fact that the delayed posttest was quite short, it is in some sense surprising that the superiority of the text structure intervention was seen on any of the measures. Perhaps if we had had the opportunity to administer a more extensive test, significant differences on the comprehension question measure would also have been noted.

These results showed that the students who were taught text structure performed significantly better both on text that had been used in the instruction and also on novel text that they had not encountered before, which is, of course, what is meant by reading comprehension. As in the previous study, there was no transfer to a less well-structured text taken from a tradebook (authentic text). That is, the comprehension of the instructed students was not significantly better than that of the students in the other treatment conditions, although the pattern of findings was as predicted. While this result is based on only one test item, it does confirm our previous findings and the general consensus that it is difficult to attain such broad transfer from an instructional program. Typically some type of bridge must be built into instruction to encourage broad transfer potential (Salomon & Perkins, 1989). Now that the feasibility of teaching text structure via this program has been demonstrated, the issue of developing the program further to encompass less well-structured text should be examined more closely.

Second grade appears to be the level at which it is feasible to begin text structure instruction. Before that point most at-risk students have such limited decoding ability that it would be difficult to provide readable texts containing sufficient cause/effect structure for effective reading comprehension instruction. The existing literature from several research traditions indicates that second graders possess a simple understanding of cause/effect and are sensitive to text structure, and it suggests that text structure instruction based on the individual sentence level would be most suitable for them. Our work has shown that an instructional focus on cause/effect as represented in single sentences can be used successfully to encourage students to begin to attend to and use structural cues (which they tend not to do; Chalmbiss, 1995; Geva, 2006), by giving them practice in accessing their rudimentary structure representations (Meyer & Ray, 2011) and helping them to move them further toward an adult representation.

Of course, these young children still have a considerable way to go (Chalmbiss & Murphy, 2002). The sentence-level effects of our intervention were maintained on the delayed test, but the paragraph-level effects were not. Transfer to authentic (i.e.,
ill-structured) text did not occur. These students will have to meet the challenges of structural variations and complexities beyond the straightforward texts used in our intervention. Given the results of our study, it appears that their progress would be enhanced by further explicit instruction in later grades.

In terms of Kintsch’s (1998) framework children at this level may need to focus on making associations between propositions within a sentence, in order to build up their representation at the local text base level, before they are able to make associations among propositions between sentences and move on to the situation model. Our intervention helps them do this, whereas less structured programs, including our Content comparison program, do not provide the support that is needed to form an adequate text base representation.

How much does the design of interventions like this one depend on developmental theory? Metz (1997) pointed out the complexity of the relation between cognitive developmental research and science curricula, a complexity that extends to other areas, including reading comprehension and social studies. Early Piagetian notions that strict developmental stages govern the type and level of thinking and reasoning have given way to an acknowledgment of considerable variability during development. According to Metz (1997), developmental theory sometimes slights the impact of experience. The knowledge that is gleaned from experience contributes substantially to one’s cognitive abilities. A consideration of children’s experience, whether resulting from formal instruction or informal learning, is, along with strictly developmental factors, of crucial importance in the design of optimal interventions. While developmental theory was not essential to the original design of this intervention (see our instructional model, described in the Introduction), findings obtained from empirical studies inspired by developmental theory were valuable to the design and helped to support and justify our design decisions.

In summary, this study indicates that the cause/effect structure can be taught successfully to second graders at risk for academic failure. Students at the second grade level have not all mastered word recognition and are not fluent readers, and they may not have a completely mature understanding of cause/effect. However, they can gain from explicit instruction in reading comprehension and should not be deprived of the opportunity to receive instruction that would provide a strong foundation for later learning.

References


(Appendix follows)
Appendix

Description of the Measures

Strategy Measures

CLUEWORDS Recalling Clue Words, oral response. The interviewer said, “What are the cause clue words?” and prompted if necessary, “What words help you find out why something happened?” Then the interviewer said, “What are the effect clue words?” and prompted if necessary, “What words help you find out what happened?” There were four such items, one for each clue word: because, since, so, therefore. The maximum score was 4.

QUESTIONS Recalling Cause/Effect Questions, oral response. The interviewer said, “What are the two cause/effect questions?” and prompted if necessary, “What two questions would you ask yourself to find the cause and the effect?” (What happened? and Why?). The maximum score was 2.

GRAPHORG Completing the Graphic Organizer, oral response. The interviewer said, “We are going to fill out this graphic organizer with the information we read in the paragraph. You tell me what to write and point to where I should write it.” There were three cause/effect sentences to be included. One point was given for each sentence that was completed correctly. The maximum score was 3.

CLAUSES Underlining Clauses in a sentence that included because and in a sentence that included therefore, written response. The interviewer said, “I am going to read you a sentence and ask you to do some things with the sentence. First, use this blue crayon to underline the cause in the sentence.” Then the interviewer said, “Use this green crayon to underline the effect in the sentence.” There were one cause and one effect clause in each sentence. Students received one point for correctly underlining the cause and the effect in each sentence. There were two such sentences. The maximum score was 2.

WEB Completing the Web Graphic Organizer, oral response. The interviewer said, “We are going to fill out this graphic organizer with the information we read in the paragraph. You tell me what to write and point to where I should write it.” There were three sentences to be included. One point was given for each sentence that was completed correctly. The maximum score was 3. (This measure was not part of the Strategy Cluster.)

Content Measures

FEATURE Feature Questions: knowledge of explicit information from the target paragraph concerning the features of the communities (home, job, school), oral response. The interviewer said, “What is a job that people had in the Cherokee community?” Answer: “They were hunters.” There were six such items. The maximum score was 6.

ADDINFO Additional Information: knowledge of explicit information from the target paragraphs, oral response. The interviewer said, “In the colonial community, what did colonists ask blacksmiths to make for them?” Answer: “Pots.” There were three such items. The maximum score was 3.

VOCAB Vocabulary Definitions, oral response. The interviewer said, “I will read you a word and you tell me what it means.” There were seven such items: Community, Asi, Cherokee, Colony, Dame School, Pioneer, Sod House. The maximum score was 7.

Sentence Combination Measures

SENTPIC Sentence Combination with Pictures, oral response. The interviewer showed students two pictures, read the sentences underneath each picture, and said, “Put these two sentences together into one sentence that explains what happened and why it happened.” Example: “The glass broke. The man dropped the glass.” One possible answer: “The glass broke because the man dropped the glass.” One point was given for correctly combining the two sentences with a clue word. There were two such items. The maximum score was 1.

SENTLISTEN Sentence Combination, Listening, oral response. The interviewer read two sentences and said, “Put these two sentences together into one sentence that explains what happened and why it happened.” Example: “It was dark. She turned on the light.” One possible answer: “It was dark, so she turned on the light.” One point was given for correctly combining the two sentences with a clue word. There were two such items. The maximum score was 2.

SENTINDEP Sentence Combination, Independent Reading, written response. The interviewer asked students to read two sentences and then said, “Put these two sentences together into one sentence that explains what happened and why it happened.” Example: “Anna drew the best picture. She won the prize.” One possible answer: “Anna drew the best picture so she won the prize.” One point was given for correctly combining the two sentences with a clue word. There were two such items. The maximum score was 2.

Comprehension Question Measures

There were four comprehension question measures. The first one assessed performance on paragraphs that had been used in instruction. The other three measures evaluated transfer. For each of the paragraphs used in these measures, there were three questions: one cause question (e.g., “Why did children write with lead?”), one effect question (e.g., “What happened because girls needed to learn how to cook and clean the house?”), and one noncausal question (e.g., “What did pioneers make out of wood?”).

(Appendix continues)
EXPLICIT Explicit Teaching. This measure was based on two paragraphs that contained information that the students had been taught explicitly. However, the questions had not been asked prior to the test. The maximum score was 6 for the two paragraphs, one point for each of the three question types in each paragraph.

(a) The interviewer read the paragraph (Colonists and their Schools), with students following along with their own copy, and said, “I am going to ask you some questions about the paragraph we just read. Be sure to answer in complete sentences.” An oral response was required.

(b) The interviewer said, “Read the paragraph (Cherokee and their Jobs) to yourself silently. Then, using the information from the paragraph, answer the questions on the worksheet.” A written response was required.

TRFR-ORAL Oral Transfer. This measure was based on two paragraphs that contained information that the students had not been exposed to in instruction; oral responses were required. The maximum score was 6 for the two paragraphs, one point for each of the three question types in each paragraph.

(a) The paragraph (Pioneers and their Games) involved a community that had been studied during the instruction and a feature that had not been presented.

(b) The paragraph (Pandas and their Habits) involved content unrelated to the content of the instructional program.

TRFR-WRIT Written Transfer. This measure was based on two paragraphs that contained information that the students had not been exposed to in instruction; written responses were required. The maximum score was 6 for the two paragraphs, one point for each of the three question types in each paragraph.

(a) The paragraph (Cherokee and their Celebrations) involved a community that had been studied during the instruction and a feature that had not been presented.

(b) The paragraph (Cities and their Subways) involved content unrelated to the content of the instructional program.

TRFR-AUTH Authentic Transfer. This measure was based on one paragraph that contained information that the students had not been exposed to in instruction. A written response was required. The paragraph (Salt and its Uses) was taken from a tradebook for children. The maximum score was 3, one point for each of the three question types.

Delayed Posttest

RECOGTOP Recognition of the Topic. This was a multiple-choice measure that included two questions. Oral responses were required. Each question presented six alternatives and asked for three choices. First, the interviewer said, “When you were in the second grade, which of these communities did you learn about in your social studies class?” One point was given for each correct answer. Then the interviewer asked, “Which of these community features did you learn about?” One point was given for each correct answer. The maximum score was 6 for the two questions.