Read Writ (2018) 31:1937–1968
https://doi.org/10.1007/s11145-018-9871-4

Comparative signaling generated for expository texts by 4th–8th graders: variations by text structure strategy instruction, comprehension skill, and signal word

Bonnie J. F. Meyer[1] · **Kausalai Wijekumar**[2] · **Puiwa Lei**[3]

Published online: 5 June 2018

Abstract Explicit signals of important relationships in expository texts can provide efficient processing instructions for readers with strategic knowledge about text structures. However, such signal words do not help readers without strategic knowledge about use of text structures and signal words. This study provided the first detailed investigation about the effects of structure strategy instruction on understanding several types of comparative signal words in multi-paragraph expository texts. The study, set in 41 school districts, examined four comparative signal words generated by three groups of reading comprehenders in Grades 4, 5, 7, and 8 and how such understandings were impacted by instruction with the text structure strategy. Students in classrooms randomly assigned to structure strategy instruction showed more understanding of comparative signal words than those in the business as usual control. The intervention aided 4th, 5th, and 7th graders' generation of all signal words, but more so for the more difficult signaling words that transitioned between paragraphs. For Grade 4 the intervention helped some reading comprehension groups more than others depending on signal word difficulty. For Grade 8 the intervention increased understanding of difficult signal words, but not the easiest signal word. Males in Grade 5 using the web-based structure strategy instruction improved their generation of the easiest signal word more than females, but females improved more on the difficult signal words. The comparison text structure and its signaling words appear ideal targets for instruction at upper elementary and middle school. The findings have implications for classroom instruction about text structures.

✉ Bonnie J. F. Meyer
bjm8@psu.edu

[1] The Pennsylvania State University, 205 Cedar Building, University Park, PA 16802-3109, USA

[2] College of Education and Human Development, Texas A&M University, College Station, TX 77843, USA

[3] The Pennsylvania State University, 140 Cedar Building, University Park, PA 16802-3109, USA

Keywords Signaling words · Rhetorical devices · Text structure strategy · Reading comprehension

Introduction

Explicit signals of important relationships within and among paragraphs in expository texts provide efficient processing instructions for readers with strategic knowledge about text structures (Meyer & Poon, 2001; Sanchez, Garcia, & Bustos, 2016). Signaling for text structures, such as "however" and "solution," can help readers notice and understand important relationships among ideas in text (Meyer, 1975). For example, signal words of "unlike" and "different" can cue readers into the comparison text structure for contrasting important ideas between two adjacent paragraphs of a text. Reader and text factors affect reading comprehension (e.g., Crossley, Rose, Danekes, Rose, & McNamara, 2017; Meyer & Rice, 1989). The reader factor of competence using the text structure strategy and the text factor of signaling provide additive effects in reading comprehension—with the greater boost coming from knowledge about strategically using text structures (Meyer & Poon, 2001). Little is known about the interaction between these particular reader and text variables with upper elementary and middle school students.

Reading comprehension and the text structure strategy

Key predictors of reading comprehension among K-12 readers include age and awareness of text structure for strategically building coherent representations and understanding nonfiction (Cain & Nash, 2011; Ray & Meyer, 2011). Also, knowledge of signaling words accounts for significant variance in reading comprehension after controlling for other reader variables, such as general vocabulary, prior knowledge, working memory, decoding, and/or rhetorical competence (Crosson & Lesaux, 2013; Sanchez et al., 2016). Furthermore, knowledge of signaling words and strategic use of that knowledge are important for understanding nonfiction (Welie, Schoonen, Kuiken, & van den Bergh, 2016). This research is important because reading comprehension of expository texts is critical for learners across the schooling years and beyond. Most K-12 students have trouble reading nonfiction texts (Hebert, Bohaty, Nelson, & Brown, 2016). Fortunately, learners from primary grades through retirement years can be taught to strategically use text structures and affordances provided by signal words (e.g., Hebert et al., 2016; Meyer & Poon, 2001; Meyer et al., 2002; Meyer & Ray, 2011; Pyle et al., 2017; Wijekumar et al. 2014; Wijekumar, Meyer, & Lei, 2017; Williams et al., 2016).

In Hebert et al.'s (2016) meta-analysis of text structure instruction, the comparison: compare/contrast text structure was the most frequently targeted text structure for instruction in interventions teaching only one or two structures of five common structures organizing nonfiction (Meyer, 1985). Hebert et al. queried whether the comparison text structure was selected more frequently for instruction because it was easier to learn or teach. However, research is mixed about

elementary school and college students' knowledge and use of the comparison text structure in relation to other text structures (e.g., Meyer & Freedle, 1984; Pyle et al., 2017; Richgels, McGee, Lomax, & Sheard, 1987). There are gaps in the extant literature about children's understanding of comparison signaling words in elementary and secondary grades as well as how instruction about the comparison text structure may impact understanding for varying skill levels within and across grades (Meyer, Ray, & Middlemiss, 2012). Therefore, the present study helps to fill these research gaps by examining the generation of several types of comparative signal words in large samples of 4th, 5th, 7th, and 8th graders at pretest and again at posttest after half of the students received text structure strategy instruction, while the other half received only the usual language arts instruction in randomized control interventions.

In the present study a generative task (Meyer et al., 2010), rather than a recognition task (Cain & Nash, 2011; Crosson & Lesaux, 2013; Geva, 2007), was used to measure children's deep understanding of comparative signal words. Deep understanding of comparative signal words was assumed when students generated appropriate signal words into blanks of a signaling test to make two-paragraph expository texts more readable (Meyer, Wijekumar, & Lin, 2011, Meyer et al., 2010, 2012). There was no training or practice of this signaling cloze task during the structure strategy instruction, and the open-response cloze signaling test served as a novel, near transfer task. Transfer tasks have been used to evaluate deeper understanding (e.g., Mayer, 2009).

The text structure strategy is a reading strategy aimed at increasing reading comprehension of nonfiction. Students learn how to produce a situation model for understanding an expository text by strategically using knowledge of signaling and text structures to make inferences in the process of building coherent mental representations of important ideas. These inferences about text structure, based on knowledge about signal words and strategic use of them, involve deep processing and are typical of students with good reading comprehension (Meyer, Brandt, & Bluth, 1980). Such processing matches characteristics of deeper processing and understanding (Graesser, 2007) and movement toward higher standards of coherence (van den Broek, Young, Tzeng, & Linderholm, 1999), rather than a default-list strategy often used by low comprehenders (Meyer et al., 1980).

Text structure instruction via intelligent tutoring of the structure strategy (ITSS)

The ITSS instruction was based on lessons in the first web-based text structure strategy instruction (Meyer et al., 2002). In this study 5th-grade students were randomly assigned to the structure strategy instruction with or without a human online-tutor or the usual reading activities that were part of the school's language arts instruction; instruction with the text structure strategy showed large effects 2 months after instruction. For a summary written for teachers about research and instruction with ITSS over the last 15 years see Meyer and Wijekumar (2016).

ITSS and the current investigation were developed and examined through a text structure model of comprehension (Meyer, 1975; Meyer & Rice, 1982; Wijekumar

et al., 2017). The goal of ITSS was to teach students to strategically use five text structures to construct organized memory structures for encoding, remembering, and flexibly using ideas learned from reading expository and persuasive texts. Instruction about the structure strategy in ITSS begins with comparison text structure, comparative signal words, and how the comparison structure can be used within other text structures to convey main ideas.

The current study is needed to specifically examine what is learned from ITSS about generating several key comparative signal words and for whom and when. We examined the understandings of comparative signaling words by readers with high, middle, and low reading comprehension skills for each grade before and after participation in the text structure strategy instruction. We wanted to see how understandings of four comparative signal words were affected by 7 months of text structure strategy instruction delivered once a week via ITSS.

Signaling with a global scope

The terms, signaling of text structure or signaling words (Meyer, 1975), have been used interchangeably with signal words, linking words and phrases, and explicit cohesive devices. Additionally, signal words overlap with some academic vocabulary words and reasoning words. Also, signaling words are similar to clue words, coherence markers, relational markers, textual rhetorical devices, and connectives, including Halliday and Hasan's classic work (1976) on cohesion. Regardless of the term used, these explicit markers of relationships among clauses, sentences, paragraphs, and larger sections of texts can provide efficient processing instructions for readers with strategic knowledge about how to use them (Geva, 2007; Meyer et al., 1980; Meyer & Rice, 1982, 1989; Meyer & Poon, 2001; Spooren & Sanders, 2008). The signaling words examined in the current research focus on signaling with a global scope (i.e., between two paragraphs) rather than a local scope (e.g., between clauses) (Lemarie, Lorch, Eyrolle, & Virbel, 2008). Research suggests that the comparison text structure, comparative signaling words, or adversative connectives are within the range of 4th- to 9th-grade students, but not yet completely mastered (Cain & Nash, 2011; McClure & Geva, 1983; Richgels et al., 1987; Meyer et al., 1980).

We were interested in students' ability to understand and use comparative signal words that logically interrelate main ideas in two adjacent paragraphs on the basis of contrasts or similarities. In our study students were asked to generate words for blanks on a signaling test that replaced targeted signal words at beginning, middle, and end of a two-paragraph expository text. For example, the signal word deleted from the beginning sentence of the two-paragraph text could provide global signaling, functioning like an advanced organizer for a reader (Meyer, 1975; Lemarie et al., 2008). An appropriate signal word for the first blank included "different" (see Table 1), but not a signal word with another meaning (e.g., "because," "the same as") or a content word (e.g., penguins). Cloze methods aimed at the global level have been used to study discourse coherence, use of top-level structure, or macro-level processing of texts (e.g., Bensoussan,1990; Levenston, Nir, & Blum-Kulka, 1984).

Table 1 Post-test signaling test

Form: penguin	Emperor penguins and Adelie penguins are_____ from one another. Emperor penguins are large penguins. They are the largest of all penguins and may grow to 4 feet tall. These penguins can weigh more than 90 lb. Emperor penguins display orange ear patches. They have long, yellow-orange streaked beaks in black faces. Emperor penguins feed principally on shallow water seafood. Emperor penguins live on Antarctica's pack ice _____ the large emperor penguins, Adelie penguins are_____ penguins. Adelie penguins grow only about 2 feet high. They weigh only about 11 lb. Adelie penguins have white ringed, beady, black eyes. Adelie penguins have short, feathered beaks on cute faces. Adelie penguins feed almost entirely on krill. _____the emperor penguins, Adelie penguins live on Antarctica's pack ice

Does understanding of signal words vary by skills, age, and instruction?

Few studies have examined age differences in elementary and middle school readers' knowledge of the comparison text structure and its signaling words with expository text. Research supports increased knowledge of signal words from early to middle school grades or beyond and with reading ability (e.g., Geva, 2007), but some research (e.g., Cain & Nash, 2011) suggests that signal word knowledge will increase to a maximum degree in elementary school with little growth thereafter. In a prior study Meyer et al. (2012) reported that students have greater understandings of the meaning of some comparative signal words over others. The difficulty in using some comparative signal words over others may relate to the logical, hierarchical structure of the text (Meyer, 1975). Comparative relationships that can organize the whole text or top-level structure (Meyer et al. 1980) may be easier to generate than contrasts between specific ideas that may be one of many compared issues.

The prior study by Meyer et al. (2012) found that readers at the middle level of reading comprehension had more significant increases across Grades 4, 6, and 9 in competency using comparative signaling than readers with low reading comprehension. Low comprehenders in both 4th and 9th graders could generate only a third of the comparative signal words; this finding is troubling considering the importance of learning from expository texts across upper elementary school through high school.

The earlier study (Meyer et al., 2012) was descriptive, cross-sectional, and involved one school district in an urban-centric suburb with 61% free or reduced lunch and 27% racial/ethnic minority students. The youngest students in both the prior and current studies were 4th graders. However, the current sample involved 5th-grade students rather than 6th graders and 7th and 8th graders rather than 9th-grade students. Moreover, the 6th graders in the prior study had nearly 2 years more of schooling (reading comprehension grade level: $M = 6.63$; $SD = 3.78$) than the 5th graders in the current study ($M = 5.15$; $SD = 2.97$) due to the grade difference and testing in late spring versus early fall. The pretest data of the current study examined the generalizability of the earlier Meyer et al. study (2012) by using a larger samples from two states with students in 4th, 5th, 7th, and 8th grades.

Current study

Of primary interest in the current study is the efficacy of ITSS for increasing low, middle, and high comprehenders' understanding four comparative signal words at each of the four grade levels. Vauras et al. (1994) reported that highly skilled readers gained more coherence building skills than less skilled readers, but we wondered if ITSS instruction would be particularly helpful for students high, middle, or low in reading comprehension on certain signal words but not as much on others. For example, in Grade 4, when expository text for learning begins to be more common, would high comprehenders working in ITSS compared to the control condition show minimal differences on the easiest signal word, but larger differences on the other more difficult signal words? Would low comprehenders show the opposite pattern with greater ITSS effects for generating "different," but smaller ITSS effects for more challenging signal words? Also, we looked at gender because of inconsistent ITSS by gender interactions reported in prior studies (Wijekumar, Meyer, & Lei, 2012; Wijekumar et al., 2014, 2017) and a signaling task that combined short verbal responses and a novel task. Females frequently perform better on verbal tasks than males, but males tend to perform better than females on novel tasks that do not require long written responses (Halpern, 2006).

Research questions

The current study provides the first detailed examination of four comparative signal words generated on the signaling test by three levels of comprehenders in Grades 4, 5, 7, and 8, and how such understandings are impacted by ITSS. Primarily, we focused on signal words generated on the posttest after completion of the intervention comparing ITSS with the business as usual control condition. Secondarily, we examined the pretest data for replication of the two main findings from the earlier Meyer et al. study (2012): (a) an interaction between reading comprehension level and grade on generation of comparative signaling, and (b) greater understanding of some comparative signal words over others. The specific questions for the pretest data included: Are main and interaction effects for grade and reading group replicated for total signaling scores and individual signal word measures? How competent are students with high, middle, and low reading comprehension in generating comparative signal words for each of the four fill-in blanks at Grades 4, 5, 7, and 8? Do results vary for different comparative signal words? Does the type of comparative signal word interact with reading comprehension group or gender at each grade level?

For the primary research questions, we posed four questions about whether for Grades 4, 5, 7 and/or 8, students nested in classrooms using the ITSS delivery of the structure strategy would outperform students nested in control classrooms on understanding of signal words.

1. Do students nested in classrooms across four grades (4th, 5th, 6th, and 8th) participating in ITSS show greater competence in generating signaling on the total signaling posttest than students not using ITSS? Do the ITSS effects on the

total signaling posttest vary for grade levels, sub-groups based on standardized reading comprehension scores, or gender?
2. Do students in the four grades (4th, 5th, 7th, or 8th) using ITSS delivery of the text structure strategy, as a partial substitute for the regular language arts program, outperform students in control classrooms across types of signal words? Is there an interaction between ITSS (ITSS instruction versus control) and type of signal word?
3. Is there an interaction between ITSS, reading comprehension group, and type of signal word at the earliest grade—Grade 4, when expository text first becomes prominent in the classroom?
4. Does the effect of ITSS interact with gender and type of signal word?

Method

Design

Signaling tests and a standardized reading comprehension were administered as part of pretest and posttest measures for four large randomized efficacy studies (e.g., Wijekumar et al., 2017). Students were nested in classrooms, which were clustered by schools. Classrooms were randomly assigned to use ITSS, as a substitution of part of their usual language arts program, or the business as usual control instruction. For the research questions that involved the four types of comparative signaling, signal word competence scores for each of the four blanks were nested within students within classrooms within schools.

Participants

Participants were 7125 students from 41 school districts located in PA or MI. The percent of students with free or reduced lunch was 43%; the percent of minority students in the sample was 12%.

High, middle, and low groups in reading comprehension

Using the same procedures and standardized reading comprehension test from Meyer et al. (2012), participants were classified into one of three comprehension ability groups: high, middle, and low on the basis of performance on a standardized test of reading comprehension administered at the beginning of the study (*Gray Silent Reading Test* [GSRT] Form B; Wiederholt & Blalock, 2000). Reliability coefficients alpha for the GSRT were reported at or above .97 for Grades 4–8. Within each grade, students were grouped according to z-scores calculated from students' raw scores. Students with z-scores of 0.55 or higher were placed into the high group; readers with scores between − 0.55 and 0.55 were placed into the middle group and readers with z-scores of −.55 or lower were classified as low.

Signaling measures

Overview of the signaling tests and open cloze format

The purpose of the texts in all forms of the signaling test was to communicate the differences between two species of the same animal living in the same region of the world (e.g., small continent, region of a large continent, small country). The texts might fit into biology curriculum in Grades 4–8 (e.g., ecological niche) providing differences in characteristics (e.g., diet and size) between species of the same animal (e.g., penguins) allowing them to live in the same area (e.g., Antarctica).

The signaling test contained four blanks for missing signaling words (see Table 1). The first blank in the first sentence of the text was designed to cue the comparison text structure. An appropriate signaling word for the first blank included "different," hundreds of possible misspellings of "different," derivatives, and paraphrases, but not a signaling word without a match to the targeted comparative relation and not words that were not signal words (i.e., content words or articles; e.g., Adelie, males, the). In the middle of the text in the first sentence of the second paragraph, two signaling words could be generated for two blanks located at the beginning and end of the sentence (see Table 1). Blank 2—"unlike," signaled the transition from one animal in the first paragraph to a different species of the animal in the second paragraph (e.g., Unlike; "In contrast to"). Blank 3—"smaller" signaled the contrasted relative size of the two animals (i.e., "larger" or "smaller"—depending on relative size of the creature in first paragraph to the one in the second paragraph). The size difference between creatures was emphasized by order of presentation in each paragraph and the first sentence of the second paragraph beginning with "Unlike." If a correct answer for blank 2—"unlike," provided a comparison between the animals in paragraphs one and two, then blank 3—"smaller" was scored correct for "smaller" or "small" because the contrast between animals on size was expressed in the two blanks working together in the middle of the paragraph. However, if relationship offered for blank 2 was not correct, then blank 3 was only scored correct if it showed the relative size with the signal word "small*er*" rather than the word "small." In contrast to signaling for the first and middle blanks comparing differences between the two animals, the final blank in the last sentence of the text highlighted the common issue of location compared for the two animals (e.g., The same as; Just like; Similar to).

The signaling test was designed to measure students' ability to generate comparative signaling words that were implied but not explicitly signaled (i.e., deleted) in expository text at a global level. We wanted to know if students could infer appropriate comparative relationships among main ideas and explicitly mark relationships with an appropriate signal word. Our focus was a correct comparative relationship (5 points or higher on a 7-point scale for each blank –5 points did not require exact matches, derivatives of the deleted signal word, or orderly syntax). Traditional cloze scoring emphasized verbatim scoring, matching semantically and syntactically –this influenced us to make the deleted signal word worth 7 points.

Also, we also wanted to know if students were generating signal words for the banks, but the relationship for the signaling was not a comparative relation or an

opposite comparative relation (4 to 3 points, respectively). We scored content words generated for the blanks and differentiated them according to how well they made some sense in the text (2 points scored) versus words that made little sense (1 point).

Alternate forms

Meyer (e.g., Meyer et al., 2010) prepared 14 forms for the signaling cloze test. Computerized, comprehensive lexical analyses (Schloss, 2017) along with traditional readability (grade level 6.4), word count (128 words), and text structure analyses (Meyer, 1985) pointed to the equivalence of the 14 signaling test forms. Applying the methods outlined in Schloss (2017), lexical properties were compared on orthographic neighborhood density, number of syllables, lexical decision time, log frequency of names response time, length, phonological neighborhood density, and number of phonemes. Analyses revealed no significant differences between the average values across all 14 texts ($p < .05$) for any of the lexical properties examined nor significant differences for pairwise bootstrapped t-tests for each pair of texts for each lexical property.

The monkey and penguin forms of the signaling test were used in the current study. These texts in the two forms focus on the comparison of two types of the same animal (monkeys [pretest], penguins [posttest]) on seven issues, including weight, length, appearance, diet, and where they live.

Scoring

Each response generated for a blank was scored using a competency of use scale (see Meyer et al., 2010, 2011, 2012); 1 point was give for a content word or article that made little sense (e.g., nice, the) and 2 points were given if the content word made some sense (e.g., far away).

Responses were scored higher than 2 points if they were signal words. Only 3 points were given for signal words that did not make sense in the context of the entire expository text. This included signal words for other text structures or the opposite meaning of the targeted comparative signaling word (e.g., identical in blank 1—"different"). Examples scored 3 points for blank 4–"the same as" included unlike, but, and different (opposite meaning), because (causal), and after (sequential). Signal words offered for the blanks were scored 4 points if they made some sense but varied from the targeted meaning and often involved unusual syntax.

Correct comparative signal words were credited with 5–7 points. A score of 5 points was given for a comparative signal word, regardless of grammar, that had the same meaning as the targeted signal word for the blank (e.g., unlike for blank 1—"different;" instead for blank 2—"unlike;" heavier for blank 3—"larger;" and similar or like for blank 4—"the same as"). Ideal equivalent words in meaning that conformed to rules of syntax and orthography were scored 7 points for blank 2—"Unlike" due to the wider array of strong contrastive signaling showing the contrast between paragraphs (i.e., In contrast to/with, Compared to/with, Different from). Most signal words generated for blank 2 were similar comparative signal words and scored 5 points (e.g., on the other hand, but, opposite, however, despite), regardless

of spelling and syntax. Derivatives of deleted signaling for a blank were scored 6 points (e.g., differ for blank 1—"different"). Signaling words that were identical to the deleted comparative signaling words, except for spelling, were scored 7 points. Analyses of results eliminating the distinction among scores 5–7 yielded the same results in pilot studies.

Human scoring of signaling words between two independent raters in past studies was 98%. A computerized scoring system, using algorithms based on student responses, was compared to human scoring by trained graduate students for the 4th and 5th grade samples in this study on the posttest. Scoring reliability ranged from .94 to .98 on the four blanks. For the current study all signaling data were rescored with the final version of the computerized scoring system; the final computer scoring and an expert's scoring for each blank was .99.

Reliability and validity in past studies

Empirical data collected in counterbalanced designs for three of the versions (i.e., monkey, penguins, and turtles) documented cloze and recall equivalence among the three forms, as well strong reliability (Meyer et al., 2010, 2012). Internal consistency using Cronbach alpha for the signaling competency scale (i.e., sum of two counterbalanced forms) was .82 (Meyer et al., 2012). Concerning validity, correlations between the signaling test and use of the structure strategy ranged from .54 to .60 for a sample of 5th graders from the two elementary schools in a district with high reading test scores (Meyer et al., 2011). For all 7th graders in the same district, the signaling test correlated slightly above .7 with (a) the reading comprehension subtests of state assessments and (b) standardized test scores for reading (i.e., *Stanford Achievement Test*).

Concurrent and predictive validity in current study

In the current study the individual signal word scores for each blank on the pretest and posttest were correlated with total signaling test scores ($p < .001$); all correlations were .42 or higher with a range from .42 to .89. Additionally, corrected total scores, removing the particular signal word/blank score from the total signaling test score, ranged from .21 to .71 ($p < .001$; also see Table 2). Individual signal word scores also correlated with the GSRT reading comprehension scores ($ps < .001$). Pretest words generated for the blanks correlated from .28 to .66 with the pretest standardized reading comprehension scores and from .24 to .59 with the posttest GSRT scores. Posttest words generated for the blanks correlated from .46 to .78 on the postest GSRT (see Table 2).

ITSS instruction

Students learned about comparative signal words through instruction and practice identifying signal words in content area texts, and use of signal words in constructing main idea statements and organizing and writing recalls. Web-based ITSS lessons about the comparison text structure emphasized writing summaries

Table 2 Evidence for concurrent validity of individual signal word measures from the signaling posttest

Grade	Signal word measures	r with total signaling test	r with corrected total signaling word test[a]	*r* with GSRT[b] posttest
Grade 4				
	Blank 1—different	0.58	0.31	0.56
	Blank 2—unlike	0.86	0.68	0.46
	Blank 3—smaller	0.83	0.66	0.51
	Blank 4—same as	0.73	0.55	0.51
Grade 5				
	Blank 1—different	0.56	0.35	0.60
	Blank 2—unlike	0.87	0.7	0.51
	Blank 3—smaller	0.84	0.67	0.54
	Blank 4—same as	0.75	0.58	0.55
Grade 7				
	Blank 1—different	0.49	0.29	0.71
	Blank 2—unlike	0.89	0.71	0.59
	Blank 3—smaller	0.85	0.67	0.63
	Blank 4—same as	0.7	0.52	0.65
Grade 8				
	Blank 1—different	0.42	0.24	0.78
	Blank 2—unlike	0.89	0.71	0.66
	Blank 3—smaller	0.84	0.63	0.68
	Blank 4—same as	0.69	0.5	0.71

[a]Correlation (*r*) with corrected total signaling word test with the contribution of correlated word removed from corrected totals for current study

[b]*Gray Silent Reading Test* [GSRT] Wiederholt and Blalock, (2000); Posttest Form A

and free recalls from texts with two or more ideas, people, or animals contrasted on three of more issues. In Grade 4 only the free recall tasks were deleted due to teachers' concern over too much writing with both the main idea and recall tasks. The comparison: compare and contrast text structure was the first structure taught in ITSS during the initial 12 lessons. The animated tutor modeled the structure strategy, communicated with students about signaling words, and provided definitions of the comparison text structure and patterns for writing a comparison main idea and a recall (see Table 3). Also, the comparison text structure was reviewed in subsequent lessons about other text structures and how they can nest hierarchically within each other.

Many of the comparison texts in the ITSS provided practice texts with the same or similar signaling words as those that would fit well into the blanks with scores from 5 to 7 on the signaling test. Table 3 identifies the comparison text structure and comparison signal words used in ITSS instruction. In the first three lessons, involving modeling and scaffolding, good signaling words were mentioned 7, 5, 0, and 0 times for blanks 1, 2, 3, and 4, respectively. For the texts in the next nine practice lessons, frequencies for signaling words matching strong answers for four blanks in the signaling test were 17, 14, 8, and 5 for blanks 1, 2, 3, and 4,

Table 3 Comparison text structure and signaling from ITSS lesson 3[a] with noted modeling/practice in other lessons[b]

Comparison	Signaling words used in comparison structure
Relates ideas on the basis of differences and similarities. The main idea is organized in parts that provide comparison between differences and similarities	instead; but; however; or; alternatively; whereas;
Pattern for Comparison Main Idea: _____ and _____(2 or more ideas) were compared on _____, _____, and _____. For example, Killer whales and Blue whales were compared on size, color, and life span	[a]*on the other hand*; while;
	compare(d); in comparison;
	in contrast (to); in opposition; not everyone;
	all but; despite; although; just; options;
	difference(s); differentiate; *different*(ly); unlike;
Pattern for Writing Recalls	
Sentence with comparison signaling word contrasting the two ideas. The first idea is…	more than; longer than; less than;
	have in common; similarities; similar(ly); share; resemble; the same as;
In contrast, the second idea is …	just as; just like, act like; look like;
	… (plus others you can find, e.g., opposite, versus, larger, smaller, better, darker)

[a]Animated agent, I.T., talking about the comparison structure and signaling in ITSS lesson 3, "We have gone through the Comparison Structure together; now we need to practice. Let's look again at the signaling words that give us clues that the author organized ideas with the comparison structure. Instead; but; however, …" (continues reading all listed below as students read along, paced as a signaling word turns color)

[b]Signaling words in *italics* were modeled by I.T. in using the structure strategy to write a main idea or recall with the comparison text in ITSS Lessons 1–2 or scaffolded for students in a practice text in lesson 3. Underlined words are signal words from texts in the nine practice lessons about the comparison text structure from ITSS Lessons 4–12

respectively (see Table 3 for specific signal words used in modeling and practice texts). The most emphasis in the lessons focused on overall structure of the text, which best matched blank 1—"different" and blank 2—"unlike." The least emphasis in the lessons focused on issues that were the same or similar for ideas compared, such as signaling to fit blank 4—"same as." The signaling tests were designed to assess what was taught in ITSS, but in a novel format that was not practiced in the ITSS lessons.

Procedures

Testing sessions for the GSRT followed by research-designed measures were conducted by researchers with teachers in the school auditorium or cafeteria. The participants were assessed with Form B of the GSRT to test reading comprehension. The signaling test (Form Monkey) was administered to examine generative understanding about comparative signaling on the pretest during fall testing. Next, students in classrooms randomly assigned to ITSS worked about 30 min per week for 7 months on the text structure strategy, while the business as usual control classrooms worked on the usual language arts instruction without ITSS. After conclusion of the intervention students completed the signaling test (Form Penguin)

in the spring. Posttests were conducted under the same conditions as the pretests for the same participants approximately 7 months later.

Data analysis

Multilevel modeling was used to account for the nesting data structure of students nested within classrooms within schools for research questions about total signaling test scores. For research questions that involved the four types of comparative signaling for blanks 1–4, blanks were nested within students within classrooms within schools. Due to the complexity of analyzing 4-level models, the four types of comparative signaling for blanks 1–4 were analyzed separately by grade level. Specifically, random-intercept ANOVA models were estimated with the Mixed procedure of SAS. ANOVA models were used because the research factors of interest were fully crossed and examining interactions among them would shed light on the research questions.

To address the replication and extension questions on pretest, a (3-level) random-intercept, 4 (grades) × 3 (comprehension groups) × 2 (gender) ANOVA model was conducted on total signaling pretest scores as well as a (4-level) random-intercept, 4 (blanks) × 3 (comprehension groups) × 2 (gender) ANOVA model for each of the four grade levels to explore if there were any significant interaction effects among the study factors.

To address research questions about the impact of text structure strategy instruction, a (3-level) random-intercept, 4 (grades) × 3 (comprehension groups) × 2 (gender) ANCOVA model was conducted on total signaling posttest scores with total signaling pretest scores controlled (research question 1). Furthermore, a (4-level) random-intercept, 4 (blanks) × 3 (comprehension groups) × 2 (treatment conditions) × 2 (gender) ANCOVA with pretest scores as covariate was conducted for each of the four grade levels to test whether treatment effects differ by comprehension levels, gender, and/or signaling words (research questions 2–4).

The highest-order significant interactions, after the Benjamini and Hochberg (1995) correction for false discovery rates at the familywise .05 level, were plotted and further examined. The Benjamini–Hochberg correction was conducted on all main and interactions effects within each ANOVA model to minimize the possibility of false discovery due to multiple tests.

Results

Did the pretest data replicate and extend the earlier study?

Meyer et al. (2012) found that (a) readers with middle level of reading comprehension had larger increases across Grades 4, 6, and 9 in competency using comparative signaling than readers with low reading comprehension, and (b) students had greater understandings of some comparative signal words over others. The current study examined competency using comparative signal words with Grades 4, 5, 7, and 8 and substantially larger samples from these four grade levels.

Were main effects and interactions replicated with Grades 4–8?

In the earlier Meyer et al. (2012) study found statistically significant main effects for grade and reading comprehension group as well as an interaction between the two factors on the total signaling test. The current study also showed significant main effects on the total signaling test for grade ($F[3, 199] = 125.46$, $p < .0001$) and reading comprehension group ($F[2, 6947] = 449.01$, $p < .0001$), but no significant interaction, $F(6,6919) = .57$, $p = .75$. Participants in Grade 4 in both the earlier and the current study did not show mastery of the comparative signaling words on the total signaling test. For both samples of 4th graders, levels of competency on the total signaling tests for high, middle, and low comprehenders averaged about 50, 40, and 30%, respectively. Within each grade of the current study, the high comprehension group scored higher than the middle group and both the high and middle groups outperformed the low reading comprehension group (e.g., Grade 8: High = 72%, Middle = 63%, Low = 51%, $p < .0001$). Overall signaling test scores for students in the middle levels of reading comprehension were similar to the earlier study. Both studies showed steady linear increases across grades for middle level comprehenders. However, there was no replication of the earlier study's sharp dip in total signaling scores for low comprehenders. The only nonsignificant difference on total signaling pretest scores between grades for the three reading groups was between low comprehenders in Grade 7 (48%) and Grade 8 (51%).

In the earlier Meyer et al. (2012) study there were also significant interactions between reading group and grade level for blank 2—"Unlike" and blank 4—"The same as." In the current study there were also statistically significant interactions for blank 2 ($p < .0001$) and blank 4 ($p = .0107$), but the interaction patterns did not replicate the earlier study's sharp drop to 4th-grade levels for low comprehenders in Grade 9. In the current study for blank 4, all three reading comprehension groups made their largest jumps in performance from Grade 5 to Grade 7. High and middle comprehenders demonstrated better understanding of this similarity signaling across each of the four grades, but low comprehenders did not differ between Grades 4 and 5 nor between Grades 7 and 8.

Main and interaction effects for extension questions

In this section we focus on the pretest data for the four types of comparative signaling words used in the four blanks of the signaling pretest. Types of comparative signal words, reading comprehension, and/or gender for each grade level were examined for main and interaction effects. We analyzed a model for each grade level because we anticipated samples and effects to be different by grade level given differences in curricula, educational experiences, and academic development across grades (Alexander, 2005). Main effects for each grade level analysis were found for reading comprehension group and word, but not gender (see top of Table 4). The main effect for word supported the conclusion of the earlier study (Meyer el al., 2012) that types of comparative signal words vary in difficulty for upper elementary and middle school students. "Different" in blank 1 was the easiest signal word for all grades in the current study and past study.

Table 4 Main and Interaction effects by grade for four comparative signal words (p values[a]) on pretest and posttest

Parameters	Grade level			
	4th	5th	7th	8th
Pretest				
Fixed effects				
Intercept	.0001*	.0001*	.0001*	.0001*
Reading comprehension group (based on GSRT: high, middle, low)	.0001*	.0001*	.0001*	.0001*
Gender	.0764	.8068	.8938	.1844
Word (signal word for each blank)	.0001*	.0001*	.0001*	.0001*
Reading comprehension group × gender	.4156	.6638	.6105	.6191
Reading comprehension group × word	.0001*	.6813	.0001*	.0001*
Gender × word	.8808	.4788	.0979	.9897
Reading comprehension group × gender × word	.7751	.4901	.1928	.2242
Random effects				
School variance	.0066	.0821	.0185	.1137
Class variance	.0094	.0187	.2351	.0321
Student variance	.0001	.0001	.0001	.0001
Residual variance	.0001	.0001	.0001	.0001
Posttest				
Fixed effects				
Intercept	.0001*	.0001*	.0001*	.0001*
Signaling word on pretest (covariate)	.0001*	.0001*	.0001*	.0001*
ITSS (text structure strategy intervention vs. control condition)	.0001*	.0001*	.0001*	.0303
Reading comprehension group (based on GSRT: high, middle, low)	.0001*	.0001*	.0001*	.0001*
Gender	.0164*	.0001*	.6741	.0984
Word (signal word for each blank)	.0001*	.0001*	.0001*	.0001*
ITSS × reading group	.7280	.4771	.3939	.1459
ITSS × gender	.4080	.7867	.2341	.7398
ITSS × word	.0001*	.0001*	.0001*	.0001*
Reading group × gender	.0442	.9966	.5408	.8731
Reading group × word	.0001*	.0001*	.0001*	.0001*
Gender × word	.1929	.0257	.5912	.0028*
ITSS × reading group × gender	.1518	.2516	.2341	.8457
ITSS × reading group × word	.0175*	.3402	.3278	.3767
ITSS × gender × word	.2992	.0097*	.8579	.5386
Reading group × gender × word	.7569	.8394	.4888	.0383
ITSS × reading group × gender × word	.4653	.1075	.4792	.8575
Random effects				
School variance	.2921	.4969	.2468	.2244

Table 4 continued

Parameters	Grade level			
	4th	5th	7th	8th
Class variance	.0004	.0002	.0118	.0082
Student variance	.0001	.0001	.0001	.0001
Residual variance	.0001	.0001	.0001	.0001

[a]Exact *p* values are provided. *indicates statistically significant fixed effects after Benjamini–Hochberg Correction (Benjamini & Hochberg, 1995)

Statistically significant interactions between reading comprehension group and word resulted for all grade levels except Grade 5 (see Table 4). The overall pattern for Grade 4 across the four signal words on the pretest showed that middle comprehenders looked more like high comprehenders for the targeted signal words "different" and "smaller," but more like low comprehenders for "unlike" and "the same as." High and middle comprehenders in Grade 4 exhibited good understanding for blank 1—"different" (*M*s > 5), while low comprehenders did not. Little understanding by 4th graders was shown by high, middle, and low reading groups for the other three signal words (*M*s for high comprehenders were below 2.02 on blank 2, below 3.29 for blank 3, and below 2.39 for blank 4), and particularly for middle and low comprehenders on blanks 2 (*M* = 1.43 for middle; *M* = 1.10 for low) and 4 (*M* = 1.73 for middle; *M* = 1.30 for low).

Scores of both high (*M* = 6.36) and middle (*M* = 6.36) comprehenders in Grade 7 indicated mastery for blank 1—"different." Although low comprehenders performed significantly lower on blank 1, an average of 5.71 indicated good performance as well. All other paired comparisons among each of the three reading comprehension subgroups on blanks 2–4 were statistically significant for 7th graders (*M*s = 4.08, 2.83, 2.24 for blank 2, *M*s = 4.93, 4.00, 3.57 for blank 3, *M*s = 3.58, 3.04, 2.22 for blank 4—for high, middle, and low comprehenders, respectively). Similar patterns of findings held over Grades 7 and 8, but with slightly higher performances for 8th graders. For Grade 8 there were no significant differences between high and middle comprehenders on both blanks 1 and 3. For blank 2—"unlike," little understanding was demonstrated by middle and low comprehenders across Grades 7 and 8. Also, average scores for all reading comprehension groups at all grades for blank 4—"the same as" fell short of adequate understanding (*M*s ranging from 1.30 to 4.02). These data suggest that ITSS may be helpful for these grade levels and all reading comprehension levels for blanks 2 through 4.

Impact of text structure strategy instruction (ITSS) on generating signal words

We addressed four questions about whether classrooms using the ITSS delivery of the text structure strategy, as a partial substitute for the standard language arts curriculum, outperformed control classrooms on understanding of signal words. The analysis conducted to answer the first intervention question examined total signaling

posttest scores and showed main effects for the ITSS intervention over the control condition (F[1, 520] = 115.43, $p < .0001$), grade (F[3, 198] = 112.56, $p < .0001$), reading comprehension groups (F[2, 7042] = 202.54, $p < .0001$), and gender (F[1, 1417] = 12.31, $p < .0005$), but no significant interactions after Benjamini–Hochberg correction (e.g., ITSS × grade F[3, 447] = 2.69, $p = .046$]). Students in the upper elementary schools and middle schools who used the ITSS delivery of the text structure strategy outperformed students in the control classrooms on the generation of signaling words. Students in higher grades scored better on the total signaling posttest than students in lower grades ($p = .014$ between Grades 7 and 8; all other comparisons ps $< .0001$). Higher reading comprehension groups scored higher than lower reading comprehension groups (ps $< .0001$), and females scored higher than males ($p = .0005$).

ITSS effects were investigated more closely by examining outcomes across the four types of signal words for each grade level. To answer research questions 2–4 we conducted a (4-level) random-intercept, 4 (signal word for each blank) × 3 (comprehension groups) × 2 (treatment conditions –ITSS vs. control) × 2 (gender) ANCOVA with pretest scores as the covariate. We ran the same analysis for each of the four grade levels to test whether ITSS treatment effects differ by comprehension levels, gender, and/or signal word (see bottom half of Table 4).

Grade 4

Grade 4 is a critical grade due to more frequent use of expository text for learning in 4th-grade classrooms than earlier elementary grades. Research question 3 focused on our expectation of an interaction between ITSS, reading comprehension group, and word in the earliest grade examined. As shown in Table 4 with 4th graders, all main effects were statistically significant. For Grade 4 the ITSS text structure intervention was effective in helping students generate better comparative signal words compared to the control condition (F[1, 91.7] = 49.01, $p < .0001$), and students performed better on blank 1–"different" than the other signal word blanks on the signaling posttest, F[3, 6033] = 736.58, $p < .0001$. Additionally, females performed better than males in generating signal words (F[1, 1861] = 5.77, $p = .016$); gender did not interact with word (F[3, 57766] = 1.58, $p = .193$) or other factors after Benjamini–Hochberg correction (see bottom of Table 4). There were three statistically significant interactions and they answered research questions 2–4. Statistically significant interactions were found for ITSS by word (F[3, 5766] = 25.20, $p < .0001$), reading comprehension group by word (F[6, 5769] = 5.33, $p < .0001$), and ITSS by reading comprehension by word (F[6, 5769] = 2.54, $p = .0175$; see Fig. 1). The ITSS by word interaction showed that ITSS boosted the generation of better responses to the four targeted signal words compared to the business as usual condition, but particularly for blanks 2 and 3. The reading comprehension group by word interaction showed that differences between high and middle reading groups were smaller on blank 1—"different" than differences between these two reading groups on the harder blanks 2, 3, and 4.

For the 3-way interaction (ITSS × reading comprehension × word) there were greater differences between ITSS and the control for certain signal words than

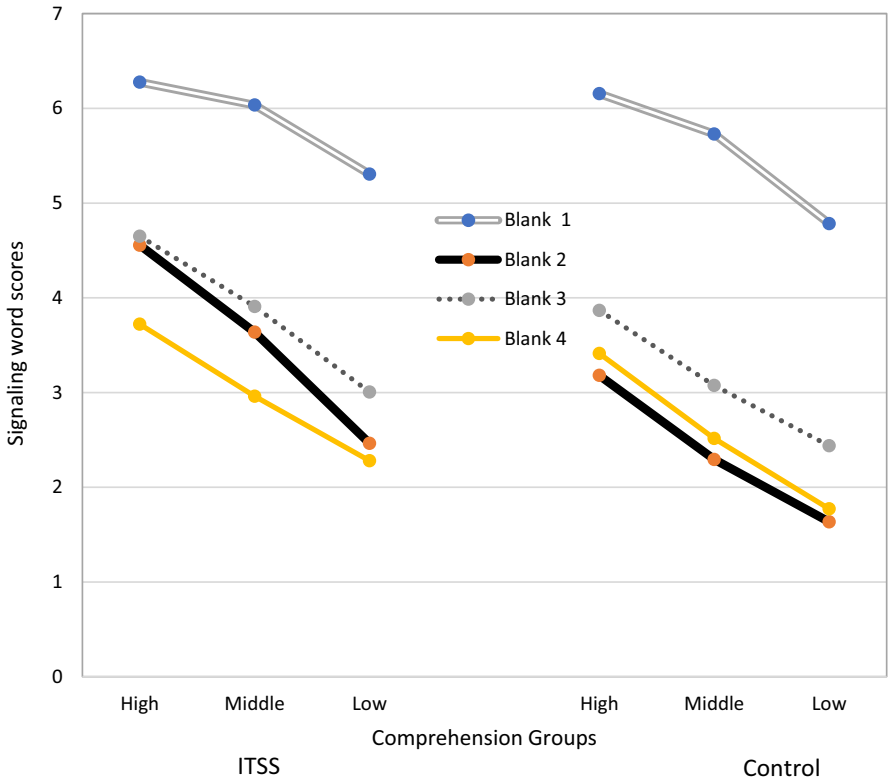

Fig. 1 Interaction among ITSS by reading comprehension group by signal word blanks on the posttest for Grade 4 (blank 1—"different," blank 2—"unlike," blank 3—"smaller," blank 4—"the same as")

others, which also varied among the reading comprehension groups. More specifically, there was no significant difference between the ITSS and the control condition for the high reading comprehension group for easiest blank (different; $p = .498$). In contrast, middle ($p = .008$) and low comprehenders ($p = .006$) performed significantly better after ITSS instruction than the business as usual controls on blank 1—"different." This same pattern held among the reading comprehension groups for blank 4—"The same as," but with much lower levels of understanding (see Fig. 1). Contrary to expectations for blank 4, there was no statistically significant difference for the high comprehenders between the ITSS and the control conditions on blank 4 ($p = .083$), but there were significant differences between ITSS and control conditions for middle ($p = .008$) and low comprehenders ($p = .006$). As seen in Fig. 1, the pattern changed with statistically significant differences ($ps < .0001$) between ITSS and the control conditions for all three reading comprehension groups on blanks 2 (Unlike) and 3 (smaller).

Grade 4: exploratory findings

The patterns in the data expected for the three reading groups were found on the easier first blank and two harder signaling words in the transition sentence introducing the second paragraph, but not for the high comprehension group on blank 4—"The same as" (see Fig. 1 blanks 2–4 for ITSS across comprehension groups). We began an exploratory investigation contrasting the pretest and posttest words generated by 4th graders for blank 4 to identify jumps in performance characterizing growth or confusions. We identified 171 students who jumped from generating a content word on the pretest to generating a comparison signal word on the posttest, but with a meaning opposite to the targeted-signal word (scored 3 points, e.g., unlike, in contrast, but). Most of these students, who generated a signal word with the opposite meaning, participated in ITSS (79%). Also, 58% of high comprehenders, who jumped from a content word on the pretest to the opposite signal word on the posttest, participated in ITSS. Some of these students may have concluded from the initial contrastive signal word that all of the rest of the blanks were to be filled with signals for the contrast between the two penguins. Alternatively, some students may have thought that any comparative signaling would work, rather than a content word, regardless of whether it signaled a difference or similarity.

For Table 5 we entered the percentage of students in Grade 4 within each treatment by comprehension cell who fell into one of three scoring categories for blank 4 on the posttest: no signal word, signal word with opposite meaning, and good signal word (see Table 5 rows 1 and 2). Column 3 in Table 5 shows that there were more high and middle comprehenders who generated content words on the posttest from the control condition than those who participated in ITSS. Column 4 in Table 5 shows that about twice as many students who generated an opposite signal word in blank 4 participated in ITSS than the control for all three abilitiy groups. Column 5 in Table 5 for generating a good signal word is similar for high comprehenders (42.3% scored 5–7) from ITSS and the control (41.2%). Students in ITSS appeared to be learning a number of strong contrastive signal words, but perhaps the students (a) did not understand the same or similar types of comparative signals, (b) had a false sense of understanding comparative signal words if they quickly located a contrastive signal word early in a text, or (c) avoided carefully reading the text on the signaling test.

Most students who generated a good signal word for blank 4 were generating one worth 5 points, a good synonym for the targeted "The same are," which signaled the same continent for the habitat of the two penguins. (The Emperor and Adelie penguins are the only two out of about 17 species of penguins that live year around on Antarctica.) The rarity for 7-point scores (i.e., the same as) for good signal words used by high comprehenders did not explain the three-way interaction; increasing scores for good synonyms from 5-point scores to 7 points did not alter the pattern of the interaction. The 3-points scores for opposite signal words appear to have particularly lowered the averages for blank 4.

Another possible reason for these findings for blank 4 relates to the deletion of recall tasks in ITSS for only Grade 4 due to concerns of teachers about too much

Table 5 Percentages of students from grade 4 and grade 8 in the six intervention × reading group cells who scored 1, 3, or at least 5 points on signal word competency for blank 4—"the same as" on the posttest[a]

Reading group	Condition	Not a signal word (e.g., for content words: Pretty, Fish, Emperor, Kings, Fat, Boys, "Luve") (%)	Signal word with meaning the opposite of the targeted signal (e.g., different from, Unlike, Not like, In contrast) (%)	Signal word matched the meaning of the targeted signal (e.g., Like, Similar to, Just like, The same as, But)
Grade 4: Comparing competency scores for no signal word, opposite signal word, and good signal word				
High	ITSS	12.2[b, a(a)]	24.6[a(b)]	37.1% (42.3%)[a(c)]
High	Control	22.4	13	35.1% (41.2%)
Middle	ITSS	27.2	28	20.9% (25.2%)
Middle	Control	36.7	14.7	19.6% (20.1%)
Low	ITSS	46.5	25	10.2% (11.6%)
Low	Control	59	11	6.3% (6.6%)
Grade 8		Not a signal word (%)	Opposite signal word (%)	Good signal word
High	ITSS	.9	21.2	68.9% (70.8%)
High	Control	4.7	9.9	76.5% (78.5%)
Middle	ITSS	6.2	27.2	53% (57.3%)
Middle	Control	10.7	15.5	52.9% (56.9%)
Low	ITSS	16.1	28	32.2% (30.6%)
Low	Control	27	24.5	27.1% (29.6%)

[a]Students' generated signal words were scored for Blank 4—"The same as" into three groups based on the signaling test competency scale: (a) not a signal word or content word with a poor match to the text (1 point); (b) signal word with meaning the opposite or incompatible with the meaning of the targeted signal word for blank 4 (3 points), and (c) good signal word of 5 points (percentages in parentheses include all individuals generating signal words scored 5–7 points)

[b]For example, this cell shows the percentage 4th-grade high comprehenders in the ITSS condition who generated a content word on the posttest for blank 4

typing. However, the deletion of the recall practice tasks in ITSS for only 4th graders is not the entire explanation of the disproportionate generation of an opposite signal word by ITSS students for blank 4. Generation of an opposite signal word for blank 4 resurfaced throughout the grades for some low comprehenders. In addition as seen in the lower half of Table 5, the pattern was also seen with the high comprehenders in ITSS in Grade 8, where ITSS effects were not found for blank 4. For Grade 8 motivation for careful reading the signaling test through the last sentence might have been lacking (e.g., Lenters, 2006). The cause for an inappropriate opposing signal words could be the instruction, a phase in learning comparative signal words, a confusion remedied with more guided practice, or lack of interest and close reading.

Grade 5

Similar to the findings for Grade 4, all main effects for Grade 5 were statistically significant as can be seen in Table 4. For Grade 5 the ITSS text structure

intervention was effective in helping students generate better signal words than the control condition ($F[1, 88.2] = 51.43$, $p < .0001$), and students performed better on blank 1–"different" than the other posttest signal words, $F[3, 6221] = 641.47$, $p < .0001$. Also, gender was statistically significant, $F[1, 1932] = 15.61$, $p < .0001$. There were three statistically significant interactions for Grade 5 including a reading comprehension group by word interaction, $F[6, 5964] = 9.05$, $p < .0001$. On average high and middle comprehenders showed minimal differences on blank 1, but there were significant differences among reading groups for the other three blanks.

The two other statistically significant interactions for Grade 5 were an interaction between ITSS and word ($F[3, 5964] = 39.43$, $p < .0001$) and a three-way interaction for ITSS by word by gender ($F[3, 5964] = 3.81$, $p = .0097$). ITSS boosted the generation of better responses to the four targeted signal words compared to the business as usual condition, but particularly on blanks 2 and 3. ITSS helped male students more than female students on blank 1–"different." However, for the other more difficult blanks 2–4, effects of ITSS were larger for females than males in generation of signal words. For females in Grade 5 the ITSS intervention was effective for the more difficult signal words, but not for the easier blank 1—"different" where 5th-grade females scored near mastery. On the other hand, males participating in ITSS generated better signal words for blank 1 as well as the harder blanks than males in the control condition. Data from the 5th graders showed that differential needs could be met by ITSS for males and females on the different types of comparative signal words.

Grade 7

Similar to findings for the earlier grades, the main effects of ITSS, reading comprehenion group, and word were statistically significant in Grade 7 (see Table 4). For Grade 7 the ITSS text structure intervention was effective in helping students generate better signal words than the control condition ($F[1, 61] = 27.19$, $p < .0001$). Also, higher reading comprehension groups generated more comparative signal words than lower reading groups, $F[2, 1726] = 49.73$, $p < .0001$. In addition, students performed better on blank 1–"different" than the other posttest signal words, $F[3, 5360] = 303.97$, $p < .0001$. However in contrast to the earlier grades, in Grade 7 gender was not statistically significant, $F[1, 256] = .18$, $p = .67$. The only statistically significant interactions were ITSS by word ($F[3, 5201] = 13.22$, $p < .0001$) and reading comprehension group by word ($F[6, 5203] = 8.15$, $p < .0001$).

ITSS participation yielded higher scores for each of the signal words than the control condition, but particularly for blanks 2 (Unlike) and 3 (smaller). The differences between 7th graders using ITSS and those in the control were statistically significant for each signal word after Benjamini–Hochberg correction. The comprehension group by word interaction for 7th graders showed smaller differences across groups of comprehenders on the first signal word ("different") than those on the other three signal words. Differences among high, middle, and low reading comprehension groups were statistically significant for all comparisons

except for the difference between high and middle comprehension groups on the first and easiest word, blank 1—"different" ($p = .24$).

Grade 8

As seen in Table 4 for Grade 8, two main effects and three interactions were statistically significant. Overall higher reading comprehension groups generated more signal words than lower groups ($F[2, 1394] = 51.51$, $p < .0001$), and students performed better on blank 1 than the other posttest signal words, $F(3, 4318) = 308.02$, $p < .0001$. The ITSS intervention effect was not statistically significant after Benjamini–Hochberg correction, $F(1, 176.3) = 4.87$, $p = .0303$. The ITSS effect, however, depended on the type of signal word as evidenced by the ITSS by word interaction ($F[2, 1394] = 51.51$, $p = .0001$). The differences between the ITSS and the control condition were statistically significant for blanks 2 and 3 (blank 2—"Unlike:" $p < .0001$; blank 3—"smaller:" $p < .0001$), but not for the easier first blank (blank 1—"different:" $p = .81$) or harder fourth blank (blank 4—"The same as:" $p = .98$). Near the end of 8th grade the first blank was near mastery levels and averaged at 6.51, 6.50, and 6.29 for high, middle and low comprehenders, respectively; 95% of the 8th graders on the posttest generated the signal word "different" or its misspelling for blank 1. However, this high level of understanding was not apparent for the 8th graders on the final signal word, "the same as."

The same analysis presented for Grade 4 in Table 5 is also presented in bottom half of the Table 5 for Grade 8. The overall higher competence level of Grade 8 students over the younger students is apparent in Table 5. Grade 8 students generated fewer content words as signals and more good signal words than Grade 4 students. The Grade 4 pattern (all reading groups of ITSS participants generated twice as many opposite comparative signal words as those in the control) was found in Grade 8, but only for only high and middle comprehenders. In the control group 24.5% of the 8th-grade low comprehenders generated opposite signal words in blank 4, which is similar to the 28% of low comprehenders in ITSS. The findings across 4th and 8th grades suggest that deleting the recall tasks for 4th graders may not be the sole contributing factor for more opposite signal words with ITSS in Grade 4. There may be an interaction between ITSS instruction and developmental and/or processing limitations.

For Grade 8 there was an interaction between gender and word ($F[3, 4198] = 4.70$, $p = .003$), but not a treatment × gender × word interaction *($F[3, 4199] = .72$, $p = .549$)*. Females did not differ from males for blank 1 ($p = .89$), blank 2 ($p = .025$—not significant after correction), and blank 4 ($p = .93$), but there was a statistically significant gender effect on blank 3 "smaller" ($p = .0125$). Females on average scored above 5 for blank 3 ($M = 5.26$), while males on average scored below 5 points ($M = 4.82$). Both females and males in Grade 8 more frequently paired "unlike" in blank 2 with "small" rather than "smaller" in blank 3; however, 45% of females showed greater use of "unlike" than males (40%). Also, females (26%) generated "smaller" more frequently than by males (22%). The current study's findings support prior research showing that females tend to exhibit better recall of verbal details than males (Halpern, 2006).

The final statistically significant interaction for Grade 8 was between reading comprehension group and word ($F[6, 4202] = 16.60$, $p < .0001$). In Grade 8 there were no differences among high ($M = 6.51$), middle ($M = 6.50$), and low ($M = 6.29$) groups in reading comprehension for blank 1 (high vs. middle: $p = .98$; high vs. low: $p = .19$; middle vs. low: $p = .18$). However, high comprehenders generated better signal words than middle comprehenders on blanks 2 and 3 for the transition between the two paragraphs of the text ($ps < .002$), but not for blank 4—"The same as" ($p = .06$). Both high and middle comprehenders generated more comparative signal words than low comprehenders for blanks 2–4 ($ps < .001$). The patterns across the four signal words varied between the high comprehension group as compared to the middle and low groups. All comprehension groups in Grade 8 generated their highest scores for blank 1 and lowest scores for blank 4—"The same as" (i.e., averaging on blank 4 at 4.47, 4.19, and 3.21 for high, middle, and low comprehenders, respectively). High comprehenders scores for blank 2–"unlike" and blank 3—"smaller" did not differ significantly ($p = . 338$). However, middle and low comprehenders performed lower on blank 2 than blank 3 ($ps < .0014$). This may indicate that (a) more high comprehenders generated excellent comparative signal words for blank 2—"unlike" than the less proficient lower two reading comprehension groups or (b) that middle and low comprehenders were better at using the signal word "smaller" than "unlike."

The former explanation above (a) best fits the data. "Unlike" was generated and paired more often with "small" than the signal word "smaller." Specifically, 52% of the high comprehenders generated "Unlike" in blank 2 and also generated small (156 cases), tiny (7 cases), or smaller (59 cases) in blank 3. In contrast, 39% of the middle comprehenders wrote "Unlike" in blank 2 and also wrote "small" (174 cases) or" smaller" (78 cases) in blank 3. Only 29% of low comprehenders generated "Unlike" in blank 2 and also paired it with small (54 cases) or smaller (29 cases) in blank 3. Most of the time low comprehenders provided a wide array of content words for blank 2—"unlike." For example, 2.5% of the low comprehenders paired "Emperor" in blank 2 with "small" in blank 3. Across the three reading groups, blank 2 was filled with the content word "Emperor" by 1% of high and middle comprehenders, but 4% of low comprehenders. Additionally, "small" was generated by 61, 58, and 51% of high, middle, and low comprehenders, respectively. The comparative signal word "smaller" was generated by 23, 27, and 24% of high, middle, and low comprehenders, respectively. High comprehenders were more proficient coordinating the meaning of the comparative relations at the beginning and end of the transition sentence between paragraphs than lower comprehenders. On the posttest 54% of low comprehenders had trouble generating a good comparative signal word in blank 2. However, some of these low comprehenders could compensate on blank 3 with their use of the signal word "smaller" to adequately generate signaling for the transition between the two paragraphs on the signaling test. For this low comprehension sub-group in Grade 8, twice as many females as males generated "smaller" for blank 3 when they did not generate a good comparative signal word for blank 2.

Discussion

This study provided the first detailed investigation about the effects of structure strategy instruction on understanding several types of comparative signal words in multi-paragraph expository texts. The primary goals for the study were to examine four comparative signal words generated by three levels of reading comprehenders in Grades 4, 5, 7, and 8 and how such understandings were impacted by instruction with the ITSS.

Summary and discussion of findings from the pretest

The pretest data replicated the positive effects in the prior study (Meyer et al., 2012) of grade and reading comprehension on generating comparative signaling, but not the interaction of these factors. Instead, the current study found a more positive trend for low comprehenders with advancing grade levels. Findings from both studies suggest that the biggest jump for low comprehenders may come during 6th grade. Both studies also showed that signal word competency varied with the type of comparative signal word tested. Like the earlier study, we found the first blank (different) on the signaling test was the least challenging, and blanks 2–4 (unlike, smaller, and the same as) were more difficult. High and average comprehenders in 7th and 8th grades mastered the comparison "different" signal word in the initial sentence of a 2-paragraph text. This signal word focused on the overall comparison text structure for the text. The findings for the current study suggest rhetorical competence (Sanchez et al., 2016) for average to high comprehenders in middle school. Findings for the elementary school samples support research by Cain and Nash (2011) that mastery occurs for some comparative connectives for many students by age 10.

The current study did not replicate low comprehenders' dip in competence using comparative signal words between Grades 6 and 9. The only nonsignificant difference among the grades on the total signaling pretest scores in the current study was between low comprehenders in Grade 7 and Grade 8. Perhaps this reflected the beginning of the decelerating changes from the end of Grade 6 to the end of Grade 9 found in the earlier study. Differences between the current study and prior study on the first blank and total signaling test scores might be due to the SES differences between the two samples, differences in curriculum related to text structure, or compliance and motivational differences for the older 9th graders in the earlier study.

Summary and discussion of the ITSS intervention and effects on comparative signal words

Classrooms randomly assigned to ITSS delivery of the text structure strategy, as a partial substitute for the standard language arts curriculum, outperformed control classrooms on generation of comparative signal words on the posttest. We further examined whether students in each of the samples at the four grade levels using

ITSS outperformed the control condition on the four targeted signal words as well as interactions of word, ITSS, comprehension group, and gender. As expected the interaction among factors of ITSS, comprehension group, and word occurred at the earliest grade—Grade 4, when expository text first became more prominent in classwork.

Summary and discussion for Grade 4

The ITSS structure strategy intervention was effective overall for the four types of signal words, but particularly for all reading comprehension groups on the two blanks transitioning from the penguin in the first paragraph to the contrasted penguin in the second paragraph (blank 2—"Unlike") as well the relative difference in size between the two penguins (blank 3—"smaller"). High comprehenders involved with ITSS did not vary appreciably from those in the control condition on the easier first signal word, "different," or the last more difficult signal word, "the same as." In contrast, middle and low comprehenders working in ITSS generated significantly higher scores for all four blanks than the middle and low groups in the business as usual control condition. Females generated more signal words than males, but gender did not interact with other variables.

We explored the quality and type of signal words used by the 4th-grade students to better understand the predicted three-way interaction among the factors of ITSS by reading group by word. Although the significant interaction was expected for research question 3, we did not expect the reduced performance of high comprehenders on blank 4 nor the lack of ITSS effects for high comprehenders on blank 4. Exploratory work resulted in sorting Grade 4 students from the reading groups within ITSS and the control into three categories based on their performance for blank 4—"the same as." The categories were: no signal word (i.e., generated a content word, e.g., fish), signal word with the opposite meaning to the targeted signaling (e.g., unlike), and a good signal word for blank 4—"The same as." On the posttest there were fewer high and middle comprehenders in the "no signal word" category from the ITSS condition than the control condition—suggesting that the ITSS students were learning to differentiate between signal words and content words. High comprehenders from ITSS and the control were similarly represented in the good signal word category. In contrast, about twice as many students, who generated an opposite signal word to the targeted signaling for blank 4, participated in ITSS than rather than the control group. These ITSS students could generate a variety of strong contrastive signal words, but perhaps the students generating an opposite comparative signal word (a) did not understand the same or similar types of comparative signals, (b) had a false sense of understanding comparative signal words if they quickly located a contrastive signal word early in a text, or (c) did not carefully read the text in the signaling test.

A supplemental reason for these findings relates to deletion of recall tasks only for the 4th-grade students due to teachers' concerns about too much typing. Relevant to blank 4, the missed recall tasks would have practiced examining a similarity for two contrasted events or creatures as one of numerous issues compared. However, the deletion of the recall practice tasks in ITSS for only 4th

graders cannot completely explain the use of an opposite signal word in blank 4 because the same errors appeared for some low comprehender across the older grades. Moreover, Grade 8 high and middle comprehenders in ITSS also generated the opposite comparative signal word for blank 4 more than high and middle comprehenders in the control condition. It will be important to find the cause(s) of incorrect generation of opposite signal words in blank 4. The problem may result from the emphasis on contrastive signaling in the ITSS instruction, a normal developmental phase in learning about comparative signal words, a need for more guided practice, or lack of close reading on the signaling test.

Summary and discussion about Grades 5 and 7

For Grade 5 the ITSS intervention was effective in helping students generate better signal words than the control condition. Students performed better on blank 1–"different" than the other posttest signal words. Females generated better signal words than males. There were three statistically significant interactions for Grade 5 that included a reading comprehension group by word interaction, an interaction between ITSS and word, and a three-way interaction for ITSS by word by gender. On average high and middle comprehenders showed minimal differences on blank 1, but significant differences among the reading groups for the other three blanks. ITSS boosted the generation of better responses to the four targeted signal words compared to the business as usual condition, but particularly on blanks 2 and 3. Also for Grade 5, ITSS helped male students to generate better signal words than males in the control group on blank 1, while ITSS enabled female students to generate better signal words than females in the control group on blanks 2–4.

Again, for Grade 7 the ITSS structure strategy intervention yielded higher signal word scores than the control condition for all of the four types of signal words, but particularly for the more difficult signal words in the last three blanks—unlike, smaller, and the same as. High and middle comprehenders were similarly proficient on the easier first signal word, "different," but varied significantly on the other signal words. The expected interaction between ITSS, reading group, and type of signal word was found Grade 4, but not Grades 5 and 7. Contrary to Grade 5, there were no gender effects or interactions for Grade 7.

Summary for Grade 8

In contrast to the earlier grades, for Grade 8 there was no main effect for ITSS, but an interaction between ITSS and signal word. Grade 8 students understood the comparative relation (different) regardless of reading comprehension group. Although the ITSS intervention had no effects on the generation of signal words for blanks 1 and 4, ITSS increased the generation of signal words for blanks 2 and 3 compared to the control condition. Generating signal words between two paragraphs was a difficult task for 8th graders without instruction in ITSS. The ITSS instruction improved 8th graders' production of signaling over the business-as-usual condition for the transition from the first penguin to the second penguin in the text on the signaling posttest. For all the grades including Grade 8, interactions were

statistically significant between reading comprehension group and signal word. On average high and middle comprehenders showed minimal differences on blank 1 for the elementary grades. By Grade 8 there were no differences among the three reading groups on blank 1, but still significant differences among all reading groups for the other three blanks, and particularly for blanks 2 and 3.

Discussion and integration of answers to the final two research questions

Research question 3 focused on an interaction among ITSS, reading comprehension group, and type of signal word, particularly for Grade 4, but not Grades 5, 7, and 8. The findings across the four grades provided an affirmation answer to research question 3. The significant interaction for Grade 4, but not the older grades, may relate to the newness of expository text structure and signal words for 4th graders with much to learn from ITSS, yet some confusions.

Research question 4 examined the interaction of gender with the other factors in our study and was posed due to Halpern's (2006) work about gender effects as well as some discrepant findings about the interaction of ITSS with gender. Specifically, Wijekumar et al. (2012) reported more progress with ITSS for 4th-grade males than females in writing short main ideas, but Wijekumar et al. (2014) found no gender effects for 5th-graders on the GSRT. However, Wijekumar et al. (2017) found a greater ITSS effect for 7th grade females than males on extended writing with multiple issues compared. In the current study there was an ITSS by gender by signal word interaction, but only for Grade 5, indicating ITSS helped only males on the easiest blank, but helped females most on the hardest blanks. For Grade 8 there was an interaction between gender and word; females did not differ from males for blanks 1, 2, and 4, but there was statistically significant gender effect on blank 3—"smaller." Specifically, more females generated "smaller," the targeted signal word for blank 3, than males. Females have demonstrated better recall of verbal details than males (Halpern, 2006). These main or interaction effects for gender in the current study would not have been visible if only total signaling tests scores were examined rather than individual signal word scores.

Implications for classroom instruction about text structures

Text structure strategy instruction can help students better understand comparative signaling. For the oldest students ITSS increased understanding of signaling with particular signal words (i.e., Grade 8 for blanks 2 and 3). For some grades ITSS improved understanding and generation of signal words for all reading groups on all words (e.g., Grade 7). In the youngest grade ITSS increased understanding overall, but particularly on certain signal words with different groups of readers based on reading comprehension skills (i.e., Grade 4). ITSS increased understanding of comparative signal words even with only 20–30 min of instruction a week over 7 months delivered by a web-based tutor without formal integration into the regular curriculum by teachers. In addition constraints with randomization of classrooms sharing the same school curtailed the use of adjunct materials in the classroom, such

as text structure keys with signal words or posters used in earlier studies (e.g., Meyer et al., 2010).

The easily administered signaling test may help teachers in classrooms assess students' grasp of comparative signal words and where to best aim instruction. Additionally individual signal word scores predicted GSRT scores and may provide extra insight for teachers in planning text structure strategy lessons for particular students. Meyer et al. (2011) showed that scores on a signal word cloze task using classroom textbooks correlated with performance on the signaling test. Teachers may be able to use selected samples of classroom materials for formative assessment of progress learning about text structure and signal words.

Most 4th graders spelled different as "diffrent," but that was just one of hundreds of misspelled versions for this word. ITSS mainly ignored spelling, but work on spelling inside and outside of structure strategy instruction may help students' confidence in using signal words in their writing.

Another practical application would be to increase examples and practice exercises for similar/same comparative signal words. We assumed from prior studies that signal words for similarities would more familiar and less difficult than contrastive comparartive signal words, but the final signaling blank was not easy for the students. ITSS focused mainly on contrasting relations as shown in Table 3, but even some high comprehenders in Grade 4 appeared to be confused about the meanings of "unlike" and "like."

Additionally, low comprehenders for blank 4 (same as) mainly generated content words that were not sensible in the context of the text. On the posttest Grade 4 readers in the low reading comprehension group (59% in the control and 47% in ITSS) most frequently generated a content word for blank 4. However from the pretest to the posttest on blank 4, 17% of the low comprehenders in ITSS moved from the 1-point content word on the pretest to "unlike" for 3-points, a signaling word with the opposite meaning. There may be a normal progression in understanding signal words that teachers and curriculum designers need to know. It may be the students go through a stage of filling in open cloze blanks by simply using the context of the near words (e.g., amazing, cute), then the context of the sentence, and finally the context of content words in the whole text. Student need to know that signal words relate closely to the text structure organizing content words, but are not the content words themselves. Once students learn the distinction between content words and signal words their strategies for using them require practice and feedback. The current study demonstrated that instruction, practice, and feedback are not just needed for the distinction between signal words for different text structures, but also between types of comparative signal words.

Related to the need for more practice and feedback, Meyer et al. (2011) examined a more individually tailored version of ITSS to provide remediation or enrichment to better match the online needs of readers. Fifth graders in the more individualized ITSS condition made more substantial gains on scores for the total signaling test from the pretest to immediate and delayed posttests than students in standard ITSS – condition used in the current study. Individualized ITSS appeared to increase understanding of signal words due to repetition of lessons with the same signal words, but different content varying in readability.

Why teach students about the compare and contrast structure and its signal words?

In answer to the question posed by Hebert et al. (2016) about why interventions about the text structure include the comparison text structure, when only two structures are covered, is that the difficulty level of the comparison text structure and its comparative signal words are suited well to students in Grades 4–8. Some comparison signaling words were understood prior to structure strategy instruction by some children, but there was little mastery. With the comparison text structure and its signaling words near an instructional range of difficulty for these grade levels, children can witness memory effects from learning the structure strategy, which could increase use and motivation to learn and apply the structure strategy for other expository text types. The problem-and-solution text structure with a cause-and-effect structure as part of the problem is memorable (Bohn-Gettler & Kendeou, 2014), but appears to be out of the reach of most 5th graders without text structure strategy instruction (Meyer, 2003) and may not serve as an ideal introductory expository text structure. The sequence may be too easy for most students at this age as well as the description structure (McClure & Geva, 1983). Simple question-and-answer and problem-and-solution text structures, without embedded causals and with familiar content (McClure & Geva, 1983), may work well for text structure instruction in primary grades (Williams et al., 2016). Similarly, familiarity of students with content words requires special consideration for ELLs with any of the expository text structures and their signal words (Crosson & Lesaux, 2013).

Overall findings of the current investigation demonstrate the importance of text structure strategy instruction for boosting upper elementary and middle school students' understanding of comparative signal words. This is important and can help students to better read, learn, and understand expository and persuasive texts used in content domains in school and everyday reading.

Limitations and suggestions for future research

Limitations of the current study include the restriction of the investigation to one text type, two-paragraph expository texts, four types of comparative signaling words, and volunteer samples from only two states. Also grade level was confounded with cohort and not longitudinal.

In addition to addressing these limitations in further research, other studies could examine on-line processing with the signaling test and similar materials for testing generation of signaling with other text structures. An important next step would be to test the same signal words as used in this study at levels of lower importance and different orders in a text to see if the current findings are characteristic of these relations represented by "different," "unlike," "smaller"/"larger," and "the same as" or if they vary with levels in the content structure (Meyer, 1975) or content domain. These different competencies may be related to the global, hierarchical text structure, but other explanations need to be evaluated related to potential differences in syntax, lexical knowledge (e.g., Crosson & Lesaux, 2013), order, proximity, or focus (McClure & Geva, 1983).

A fruitful area of study may be stages of development for competence in understanding comparative signal words, and if the patterns and categories of change seen in the exploratory analyses hold for other participants with different backgrounds. A next step study could explore the relationship between needs for coherence (van den Broek et al., 1999) at the sentence, paragraph, or global text levels with the stages of developing competence. Additionally the "no signal word," "signal word with the opposite meaning," and "good signal word" categories could be further explored to see whether they are practical categories and relate to cognitive resources, reading comprehension, vocabulary, writing competency for main ideas, use of the structure strategy, and/or reading time.

Acknowledgements The research reported here was supported by the Institute of Education Sciences, U.S. Department of Education, through Grants R305G030072, R305A080133, and R305A120593 to The Pennsylvania State University and IES Grant R305A150057 to Texas A & M University. The opinions expressed are those of the authors and do not represent views of the Institute or the U.S. Department of Education.

References

Alexander, P. A. (2005). The path to competence: A lifespan developmental perspective on reading. *Journal of Literacy Research, 37,* 413–436. https://doi.org/10.1207/s15548430jlr3704_1.

Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society, 57*(1), 289–300.

Bensoussan, M. (1990). Redundacy and the coherence cloze. *Journal of Research in Reading, 13*(1), 18–37.

Bohn-Gettler, C. M., & Kendeou, P. (2014). The interplay of reader goals, working memory, and text structure during reading. *Contemporary Educational Psychology, 39*(3), 206–219.

Cain, K., & Nash, H. M. (2011). The influence of connectives on young readers' processing and comprehension of text. *Journal of Educational Psychology, 103,* 429–441. https://doi.org/10.1037/a0022824.

Crossley, S. A., Rose, D. F., Danekes, C., Rose, C. W., & McNamara, D. S. (2017). That noun phrase may be beneficial and this may not be: Discourse cohesion in reading and writing. *Reading and Writing, 30,* 569–589.

Crosson, A. C., & Lesaux, N. K. (2013). Does knowledge of connectives play a unique role in reading comprehension of English learners and English-only students? *Journal of Research in Reading, 36*(3), 241–260.

Geva, E. (2007). Conjunction use in school children's oral language and reading. In R. Horowitz (Ed.), *Talking texts: How speech and writing interact in school learning* (pp. 271–294). Hillsdale, NJ: Erlbaum.

Graesser, A. C. (2007). An introduction to strategic reading comprehension. In D. S. McNamara (Ed.), *Reading comprehension strategies: Theories, interventions, and technologies* (pp. 3–26). New York: Lawrence Erlbaum Associates, Taylor & Francis Group.

Halliday, M. A. K., & Hasan, R. (1976). *Cohesion in English*. London, England: Longman.

Halpern, D. F. (2006). Assessing gender gaps in learning and academic achievement. In E. Anderman, P. H. Winne, P. A. Alexander, & L. Corno (Eds.), *Handbook of educational psychology* (2nd ed., pp. 635–654). Mahwah, NJ: Lawrence Erlbaum Associates.

Hebert, M., Bohaty, J. J., Nelson, J. R., & Brown, J. A. (2016). The effects of text structure instruction on informational text comprehension: A meta-analysis. *Journal of Educational Psychology, 108,* 609–629.

Lemarié, J., Lorch, R. F., Eyrolle, H., & Virbel, J. (2008). SARA: A text-based and reader-based theory of signaling. *Educational Psychologist, 43*(1), 27–48. https://doi.org/10.1080/00461520701756321.

Lenters, K. (2006). Resistance, struggle, and the adolescent reader. *Journal of Adolescent & Adult Literacy, 50*(2), 136–146. https://doi.org/10.1598/JAAL.50.2.6.

Levenston, E. A., Nir, R., & Blum-Kulka, S. (1984). Discourse analysis and the testing of reading comprehension by cloze techniques. In A. K. Pugh & J. M. Ulijn (Eds.), *Reading for professional purposes*. London: Heinemann.

Mayer, R. E. (2009). *Multimedia learning* (2nd ed.). New York: Cambridge University Press.

McClure, E., & Geva, E. (1983). The development of the cohesive use of adversative conjunctions in discourse. *Discourse Processes, 6,* 411–432. https://doi.org/10.1080/01638538309544575.

Meyer, B. J. F. (1975). *The organization of prose and its effects on memory*. Amsterdam: North-Holland.

Meyer, B. J. F. (1985). Prose analysis: Purposes, procedures, and problems. In B. K. Britton & J. Black (Eds.), *Analyzing and understanding expository text* (p. 11-64, 269-304). Hillsdale, NJ: Erlbaum.

Meyer, B. J. F. (2003). Text coherence and readability. *Topics in Language Disorders, 23,* 204–221.

Meyer, B. J. F., Brandt, D. M., & Bluth, G. J. (1980). Use of the top-level structure in text: Key for reading comprehension of ninth-grade students. *Reading Research Quarterly, 16,* 72–103.

Meyer, B. J. F., & Freedle, R. O. (1984). Effects of discourse type on recall. *American Educational Research Journal, 21,* 121–143.

Meyer, B. J. F., Middlemiss, W., Theodorou, E., Brezinski, K. L., McDougall, J., & Bartlett, B. J. (2002). Effects of structure strategy instruction delivered to fifth-grade children via the Internet with and without the aid of older adult tutors. *Journal of Educational Psychology, 94,* 486–519.

Meyer, B. J. F., & Poon, L. W. (2001). Effects of structure strategy training and signaling on recall of text. *Journal of Educational Psychology, 93,* 141–159.

Meyer, B. J. F., & Ray, M. N. (2011). Structure strategy interventions: Increasing reading comprehension of expository text. *International Electronic Journal of Elementary Education, 4*(1), 127–152.

Meyer, B. J. F., Ray, M. N., & Middlemiss, W. (2012). Children's use of comparative text signals: the relationship between age and comprehension ability. *Discours. Revue de linguistique, psycholinguistique et informatique*, *10*, 1–25.

Meyer, B. J. F., & Rice, G. E. (1982). The interaction of reader strategies and the organization of text. *Text-Interdisciplinary Journal for the Study of Discourse, 2*(1–3), 155–192.

Meyer, B. J. F., & Rice, G. E. (1989). Prose processing in adulthood: The text, the reader, and the task. In L. W. Poon, D. C. Rubin, & B. A. Wilson (Eds.), *Everyday cognition in adulthood and later life* (pp. 157–194). New York: Cambridge University Press.

Meyer, B. J. F., & Wijekumar, K. (2016). Intelligent tutoring of the structure strategy: A reading strategy tutor. In S. A. Crossley & D. S. McNamara (Eds.), *Adaptive educational technologies for literacy instruction* (pp. 82–103). New York, NY: Routledge Publishers, Taylor & Francis Group.

Meyer, B. J. F., Wijekumar, K., & Lin, Y. (2011). Individualizing a Web-based structure strategy intervention for fifth graders' comprehension of nonfiction. *Journal of Educational Psychology, 103*(1), 140–168.

Meyer, B. J. F., Wijekumar, K., Middlemiss, W., Higley, K., Lei, P., Meier, C., et al. (2010). Web-based tutoring of the structure strategy with or without elaborated feedback or choice for fifth- and seventh-grade readers. *Reading Research Quarterly, 45*(1), 62–92.

Pyle, N., Vasquez, A. C., Lignugaris/Kraft, B., Gillam, S. L., Reutzel, D. R., Olszewski, A., et al. (2017). Effects of expository text structure interventions on comprehension: A meta-analysis. *Reading Research Quarterly, 52*(4), 469–501. https://doi.org/10.1002/rrq.179.

Ray, M. N., & Meyer, B. J. F. (2011). Individual differences in children's knowledge of expository text structures: A review of literature. *International Electronic Journal of Elementary Education, 4,* 67–82.

Richgels, D. J., McGee, L. M., Lomax, R. G., & Sheard, C. (1987). Awareness of four text structures: Effects on recall of expository text. *Reading Research Quarterly, 22,* 177–196. https://doi.org/10.2307/747664.

Sanchez, E., Garcia, J. R., & Bustos, A. (2016). Does rhetorical competence moderate the effect of rhetorical devices on the comprehension of expository texts beyond general comprehension skills? *Reading and Writing, 30*(3), 439–462. https://doi.org/10.1007/s11145-016-9684-2.

Schloss, B. J. (2017). *Predicting concept evolution during naturalistic reading with simultaneous eye-tracking and fMRI* (Unpublished master's thesis). The Pennsylvania State University, University Park, PA.

Spooren, W. P. M., & Sanders, T. (2008). The acquisition order of coherence relations: On cognitive complexity in discourse. *Journal of Pragmatics, 40,* 2003–2026. https://doi.org/10.1016/j.pragma.2008.04.021.

van den Broek, P., Young, M., Tzeng, Y., & Linderholm, T. (1999). The Landscape model of reading: Inferences and the online construction of a memory representation. In S. R. Goldman & H. van

Oostendorp (Eds.), *The construction of mental representations during reading* (pp. 71–98). Mahwah, NJ: Lawrence Erlbaum Associates.

Vauras, M., Kinnunen, R., & Kuusela, L. (1994). Development of text-processing skills in high-, average-, and low-achieving primary school children. *Journal of Reading Behavior, 264,* 361–389.

Welie, C., Schoonen, R., Kuiken, F., & van den Bergh, H. (2016). Expository text comprehension in secondary school: For which readers does knowledge of connectives contribute the most? *Journal of Research in Reading, 40,* S42–S65. https://doi.org/10.1111/1467-9817.12090.

Wiederholt, J. L., & Blalock, G. (2000). *Gray silent reading tests*. Austin, TX: Pro-Ed.

Wijekumar, K. K., Meyer, B. J. F., & Lei, P. (2012). Large-scale randomized controlled trial with 4th graders using intelligent tutoring of the structure strategy to improve nonfiction reading comprehension. *Journal of Educational Technology Research and Development, 60,* 987–1013. https://doi.org/10.1007/s11423-012-9263-4.

Wijekumar, K(. K)., Meyer, B. J. F., & Lei, P. (2017). Web-based text structure strategy instruction improves seventh graders' content area reading comprehension. *Journal of Educational Psychology, 109*(6), 741–760.

Wijekumar, K., Meyer, B. J. F., Lei, P., Lin, Y., Johnson, L. A., Spielvogel, J. A., et al. (2014). Multisite randomized controlled trial examining intelligent tutoring of structure strategy for fifth-grade. *Journal of Research on Educational Effectiveness., 7,* 331–357. https://doi.org/10.1080/19345747.2013.853333.

Williams, J. P., Kao, J. C., Pao, L. S., Ordynans, J. G., Atkins, J. G., Cheng, R., et al. (2016). Close analysis of texts with structure (CATS): An intervention to teach reading comprehension to at-risk second graders. *Journal of Educational Psychology, 108*(8), 1061–1077. https://doi.org/10.1037/edu0000117.