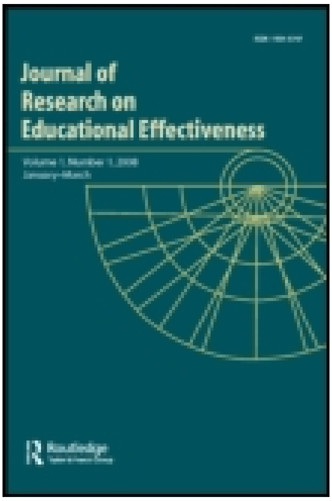

Journal of Research on Educational Effectiveness
Publication details, including instructions for authors and subscription information:
http://www.tandfonline.com/loi/uree20

Multisite Randomized Controlled Trial Examining Intelligent Tutoring of Structure Strategy for Fifth-Grade Readers

Kausalai Wijekumar[a], Bonnie J. F. Meyer[b], Pui-Wa Lei[b], Yu-Chu Lin[b], Lori A. Johnson[c], James A. Spielvogel[c], Kathryn M. Shurmatz[c], Melissa Ray[b] & Michael Cook[b]

[a] Texas A&M University, College Station, Texas, USA

[b] The Pennsylvania State University, University Park, Pennsylvania, USA

[c] The Pennsylvania State University, Monaca, Pennsylvania, USA

Accepted author version posted online: 10 Mar 2014.Published online: 29 Sep 2014.

Journal of Research on Educational Effectiveness, 7: 331–357, 2014

ISSN: 1934-5747 print / 1934-5739 online
DOI: 10.1080/19345747.2013.853333

Multisite Randomized Controlled Trial Examining Intelligent Tutoring of Structure Strategy for Fifth-Grade Readers

Kausalai Wijekumar
Texas A&M University, College Station, Texas, USA

Bonnie J. F. Meyer, Pui-Wa Lei, and Yu-Chu Lin
The Pennsylvania State University, University Park, Pennsylvania, USA

Lori A. Johnson, James A. Spielvogel, and Kathryn M. Shurmatz
The Pennsylvania State University, Monaca, Pennsylvania, USA

Melissa Ray and Michael Cook
The Pennsylvania State University, University Park, Pennsylvania, USA

Abstract: This article reports on a large scale randomized controlled trial to study the efficacy of a web-based intelligent tutoring system for the structure strategy designed to improve content area reading comprehension. The research was conducted with 128 fifth-grade classrooms within 12 school districts in rural and suburban settings. Classrooms within each school were randomly assigned to intervention or control groups. The intervention group used the intelligent web-based tutoring system for the structure strategy (ITSS) for 30 to 45 min each week as a partial substitute for the language arts curriculum for the entire school year. The structure strategy teaches students how to read and comprehend expository texts by identifying the text structure and creating strategic mental representations of the text. The web-based tutoring system delivered the structure strategy training with modeling, practice tasks, assessment, and feedback. The control classrooms used the school's language arts curriculum for the full language-arts time. Results show that the ITSS delivered structure strategy training improved reading comprehension measured by a standardized test and researcher designed measures.

Keywords: Reading comprehension instruction, web-based tutoring, RCT, efficacy, structure strategy, fifth-grade reading

Reading and comprehending textbooks, newspapers, and other informational sources are essential components of lifelong success (Biancarosa & Snow, 2006). Whether children can read, understand, make connections to prior knowledge, create flexible and accessible memory of texts, and effectively recall and use the acquired knowledge plays a significant role in their success (Alexander, 2005; Pressley, 2002). Unfortunately, national and state assessments of reading comprehension in our schools show a dismal picture with 33% of fourth graders reading below basic levels of proficiency (National Center for Education

Address correspondence to Kausalai Wijekumar, Professor of Teaching, Learning, and Culture, Texas A&M University, 420C Harrington Tower, 4232 TAMU, College Station, TX 77843, USA. E-mail: K_Wijekumar@tamu.edu

Color versions of one or more of the figures in the article can be found online at www.tandfonline.com/uree.

Statistics, 2011). The structure strategy—a text-structure-based reading comprehension approach—has shown potential in small studies at fifth grade and one large efficacy study at fourth grade (Wijekumar, Meyer, & Lei, 2012). The goal of this large-scale randomized controlled trial was to study whether the structure strategy, delivered through an intelligent web-based tutoring system (ITSS) to children in fifth grade, can improve their comprehension of expository texts.

PURPOSE OF THE STUDY

After the primary grades, students increasingly are expected to learn from expository texts in science, history, social studies, and current events (Common Core Standards Initiative, 2010; Gersten, Fuchs, Williams, & Baker, 2001; Guthrie & Davis, 2003; Wilson & Rupley, 1997). The focus of this study was on reading and comprehending expository texts, such as those found in social studies, science, and history textbooks.

Fourth and fifth grade appear to be critical times in transitioning students from narrative to expository text comprehension. The structure strategy and its delivery using ITSS were tested in a large scale efficacy study in fourth grade and showed statistically significant improvements in reading comprehension (Wijekumar, Meyer, & Lei, 2012). We selected fifth grade for this research study because of the significance of this year in a child's progress to middle school comprehension needs in Grade 6 and beyond.

Activities designed to improve reading comprehension and used in current language arts textbooks and language arts curricula include word recognition, developing vocabulary, developing and activating background knowledge about text, visualizing, highlighting important ideas, summarizing, and questioning (Cantrell, Almasi, Carter, Rintamaa, & Madden, 2010; Ponce, López, & Mayer, 2012). Curricula used in the schools participating in this study were Harcourt Storytown (4%), Harcourt Trophies (8%), Houghton Mifflin (35%), McGraw Hill Adventures in Reading (10%), Renaissance Learning Accelerated Reader (24%), Scott Foresman Reading Street (24%), and custom solutions developed by the schools (18%). The percentages do not add up to 100% because some schools used two of these curricula together. These curricula focused on background knowledge, question generation, summary writing, and some graphical representation of texts and mentioned one or two of the text structures, but did not provide the in-depth, strategic instruction about using all the text structures as delivered in the structure strategy (Meyer & Poon, 2001). In this study the intervention classrooms substituted one class period a week with structure strategy instruction via the web-based ITSS. During the other 4 days of the week the intervention classrooms used the regular language arts curriculum used in the school. The control classrooms used the regular language arts curriculum for all 5 days of the week. The multisite cluster randomized design for this study was designed to ensure that the school/cluster used the same reading programs and varied only on ITSS use.

Reading comprehension approaches, such as activating background knowledge, question generation, summary writing, creating graphical representation of texts (Cantrell et al., 2010; Ponce et al., 2012), and the structure strategy share some theoretical foci but differ in how the theory is operationalized in practice. One shared focus is that reading comprehension requires learners to actively read and understand ideas. The ideas need to be integrated with the learner's prior knowledge into effective and efficient memory structures that will be accessible for future use in problem solving, generating inferences, and other applications.

The theoretical framework for the structure strategy differs from the frameworks of some of the other approaches to reading comprehension because it developed out of empirical studies integrating and studying the effects of intuitive logical structures (Meyer

& McConkie, 1973), linguistics categories (Grimes, 1975; Halliday & Hasan, 1976), and psychological variables in the verbal learning tradition. This research found effects of text structure on identification of important information from expository text, memory representations of the text, and what readers remembered after reading (e.g., Crothers, 1972; Frederiksen, 1975; Kintsch & van Dijk, 1978; Meyer, 1975). The structure strategy facilitates understanding of text by organizing concepts based on explicit or implied relationships that communicate main ideas. Good readers are able to classify the text structure used by the authors; focus on the logical structure in the text; and take advantage of text structure to identify or construct main ideas, organize their memory, and effectively produce coherent representations of the text, when needed (Meyer, Brandt, & Bluth, 1980). Meyer et al. distinguished this strategy, called the structure strategy, from a default/list strategy in which the text is viewed and processed as a series of loosely related propositions.

The structure strategy developed through multiple research studies (see Table 1) operationalizes the theory using signaling words and main idea patterns unique to each text structure. Additional aspects of the structure strategy include connecting strategic encoding based on text structure patterns to prior knowledge, combining text structures reflective of authentic texts, and comprehension monitoring. In contrast, underlining, summary writing, questioning, and other techniques used in the schools' curricula are not as explicit as the structure strategy about what constitutes effective and efficient memory structures for encoding and retrieving important ideas from expository texts. However, these common current approaches in classrooms can be combined with the structure strategy training to improve reading comprehension. For example, highlighting important information can be improved by using text structure to find the most important ideas in the text. Our focus in this study is teaching fifth-grade children how to read and comprehend expository texts using the structure strategy.

Most of the previous work on teaching children about text structure has used trained teachers or researchers to deliver the instruction face to face. Pressley, Wharton-McDonald, Mistretta-Hampston, and Echevarria (1998) reported on literacy instruction in fourth- and fifth-grade classrooms and noted a lack of instruction in comprehension strategy. Expert teachers in the study either were not aware of comprehension strategies or were unable to deliver instruction on the strategies. ITSS was designed to address these challenges as well as the need to reach fifth-grade students from different geographical locations who might not encounter structure strategy instruction without a web-based tutor. For this study we delivered ITSS to both rural and suburban school settings.

Biancarosa and Snow (2006) recommended the use of direct and explicit strategy instruction, strategic tutoring, diverse texts, and technology as a tool for literacy instruction as elements to improve middle-school literacy. Technology based interventions can provide consistency of delivery, practice tasks, assessment, and feedback (Proctor, Dalton, & Grisham, 2007). ITSS was built using many features synthesized from previous research on automated tutors (Graesser et al., 2004), cognitive tutors (Aleven & Koedinger, 2002; Anderson, Corbett, Koedinger, & Pelletier, 1995), pedagogical agents (Baylor, 2001), multimedia learning (Mayer, 2001), and technology-based affordances (Meyer & Wijekumar, 2007). Finally, this research study focused on the efficacy of the structure strategy delivered using this technology-based infrastructure.

THE STRUCTURE STRATEGY AND WEB-BASED DELIVERY

For more than 40 years, Meyer (e.g., 1971, 1975) and her colleagues (e.g., Meyer et al., 2010) have studied text structures (e.g., comparison, problem-and-solution, cause-and-effect,

Table 1. History of structure strategy used in ITSS

Study	Focus	Lessons Learned
Meyer & McConkie (1973); Meyer (1975)	What good readers remembered from hierarchically organized logical text from *Scientific American.* Whether some text structures are more memorable than others and the interaction of reading strategies and text structure.	Logical structure of a passage is related to certain aspects of the cognitive structure that the learners construct. Text structures/top-level structures are important for characterizing the logical structure of text's main ideas as well a reader's understanding of those main ideas.
Meyer (1975)	Type and structure of relationships (Grimes, 1975) among concepts in text dramatically influence comprehension when they occur at the top third of the structure, but not low in the text structure.	Value for reading comprehension research and applications of studying *text structure as it organizes the main ideas of a text* rather than the details.
Meyer, Brandt, & Bluth (1980)	Whether reading skills of ninth graders relate to using the structure strategy. Comparing the top-level structure of a text to the top-level structure organizing a student's recall and how this relates to signaling in text (e.g., "however" for comparison text structure).	Good readers with organized text use the structure strategy regardless of signaling. Readers with low reading skills do not use the structure strategy. Student lower in reading comprehension than vocabulary (or decoding) could use the structure strategy, but only for texts with signaling. Instruction in the structure strategy is needed.
Bartlett (1978)	First direct instruction to teach structure strategy (causation, collection, comparison, description, and problem-solution) based on Meyer's work (Meyer, 1975; Meyer, Brandt, & Bluth, 1980). Students taught to find main idea first then determine text structure.	Instruction about text structure increased students' ability to identify and use the text's top-level structure and increased recall over a class receiving spelling instruction.
Englert & Thomas (1987)	Four text structures were taught (description, enumeration, sequence, and compare-contrast).	Some text structures are easier than others (e.g., sequence or listing easier than compare/contrast).
Meyer & Rice (1989)	Modified Bartlett's approach by spotlighting signaling words to identify text structure and top-level structure prior to writing main ideas. Trained instructors taught young and older adults in classroom settings. Instructors and peers modeled the use of main idea patterns unique to each text structure to compose written main ideas and recall protocols.	The structure strategy in today's form begins to take shape: 1. Search for the signaling words and top-level text structure of the text that could inter-relate all the ideas in the text. 2. Top-level structure leads the reader to the main idea of the text. 3. Assist the reader in constructing a coherent representation or situation model. 4. Main idea was identified as the ideas interrelated by the top-level structure. 5. Explicit instruction about how these text structures often combined in a wide variety of everyday texts, such as a problem with causes and three favored solutions to eliminate causes versus a solution that was not favored by the author. Effect sizes .64 for recall and .52 for main idea.

(Continued on next page)

Table 1. History of structure strategy used in ITSS *(Continued)*

Study	Focus	Lessons Learned
Meyer & Poon (2001)	Younger and older adults trained on the structure strategy with customized main idea patterns for each text structure.	Emphasis and practice in using the customized main idea patterns for each text structure. Increased total recall from a variety of texts ($d = .64$), informative video ($d = 1.47$), medical decision-making task ($d = .93$).
Meyer & Poon (2001)	Adults with slightly reduced working memory learned text structure better sequentially, singly or in pairs, rather than all 5 at once.	Introduce the five text structures sequentially starting with comparison.
Theodorou (2006)	Web-based and teacher-led training of the structure strategy – problem and solution text structure for college students.	Teaching in one domain does not transfer to other domains. Explicit instruction in many different domains necessary for transfer.
Meyer et al. (2002)	Internet-based lessons for the structure strategy for fifth graders. 2.5 months of training three times a week for 20 min a session where students silently read web page lessons and wrote main ideas and recalls on the web. Older adults tutors responded online to students with 1- to 2-day delayed feedback.	Three step approach with signaling words, identifying text structure (top-level structure), writing a main idea, and writing a recall. Sequencing of instruction about text structures (first comparison, then problem and solution, followed by cause and effect, sequence, and ended with description and collection), combining text structures. Increases in ideas recalled $2\frac{1}{2}$ months after instruction ($d = .92$).
Meyer et al. (2010)	Web-based intelligent tutoring system ITSS was used to present the structure strategy to fifth- and seventh-grade students. Two factor experiment with pre- and multiple posttest design.	56 fifth graders and 55 seventh-grade students interacted with ITSS for the school year and results showed that students receiving ITSS with elaborated feedback showed greater pretest to posttest gains on the GSRT ($d = .55$) than students receiving ITSS with simple feedback ($d = .15$). Lesson choice as a motivational element did not improve the outcomes.

Note. ITSS = intelligent web-based tutoring system for the structure strategy; GSRT = Gray Silent Reading Test.

sequence, description) in science, social studies, and other domains. Details about the development and research on the structure strategy are presented in Table 1.

Good readers use their knowledge of text structures to build coherent memory representations. Signaling words (such as "in contrast," "as a result," "solution") can cue text structure and guide readers toward coherent text representations with a key role in selection and encoding (e.g., Lorch & Lorch, 1995; Meyer & Poon, 2001). During structure strategy instruction, readers learn to identify specific signaling words for each of the five text structures covered by the structure strategy. They proceed to impose a top-level structure on the text and create a main idea using a pattern tailored to each type of text structure. This process is designed to guide the construction of strategic memory of the text. Combining

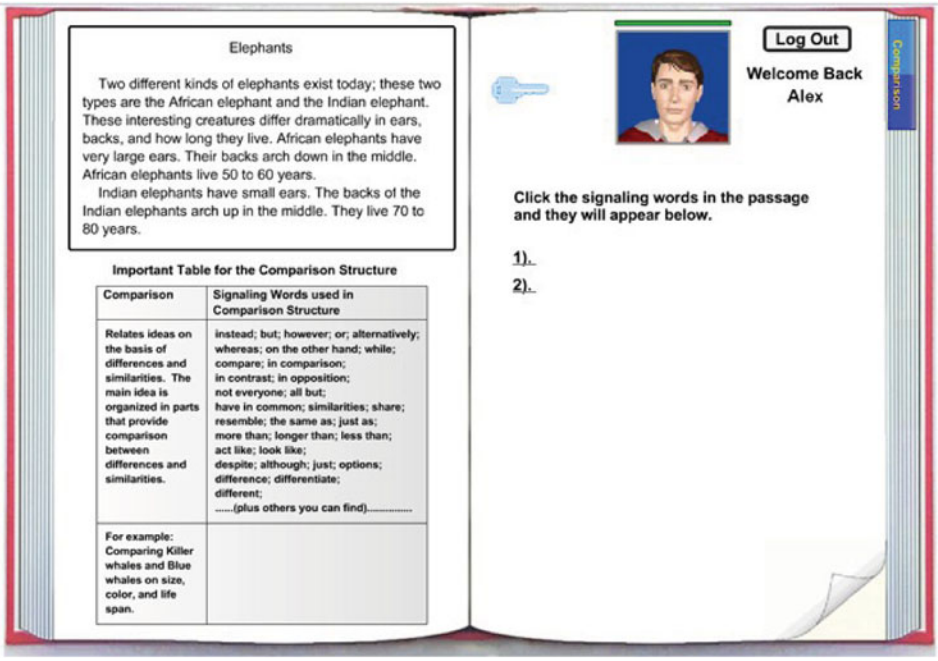

Figure 1. Intelligent tutoring system for the structure strategy book-like screen minimizing seductive details.

both the signaling words and the main idea allows the readers to facilitate comprehension monitoring of the details in a text.

ITSS is a web-based intelligent tutor developed using many previous research studies and the historical antecedents of the intervention are summarized in Table 1. Some key elements of the design summarized in Table 1 include the use of the five text structures, sequencing of the text structures, use of the specific patterns for each text structure, and presentation of the structure strategy in many different domains. The booklike interface shown in Figure 1 is designed specifically to minimize seductive details (Mayer, 2001) and emphasize the text being read and the tutoring role of the animated pedagogical agent—intelligent tutor (I.T.).

I.T. begins a lesson with a description of objectives (e.g., "Today we are going to learn about the comparison text structure") and models how he would perform the task (e.g., "When I read a passage I look for signaling words"). I.T. proceeds to read the passage to the learner, highlights the signaling words (signaling words flash in a different color on the web page), and then asks the student to practice by finding the signaling words in a similar passage (Figure 1). When the student clicks on the signaling words, I.T. displays his or her answer on the right side of the book. After the student has completed his or her answer, I.T. processes the input by parsing, checking spelling and synonyms, and then checking the answers against the database. If the answers are correct, I.T. gives positive feedback (e.g., "Excellent job!"). If the answer is partially correct, I.T. uses a series of rules built into the database that identify the most appropriate feedback to give to the student (e.g., "Please correct your signaling words"). Students then identify which text structure is used to organize the passage.

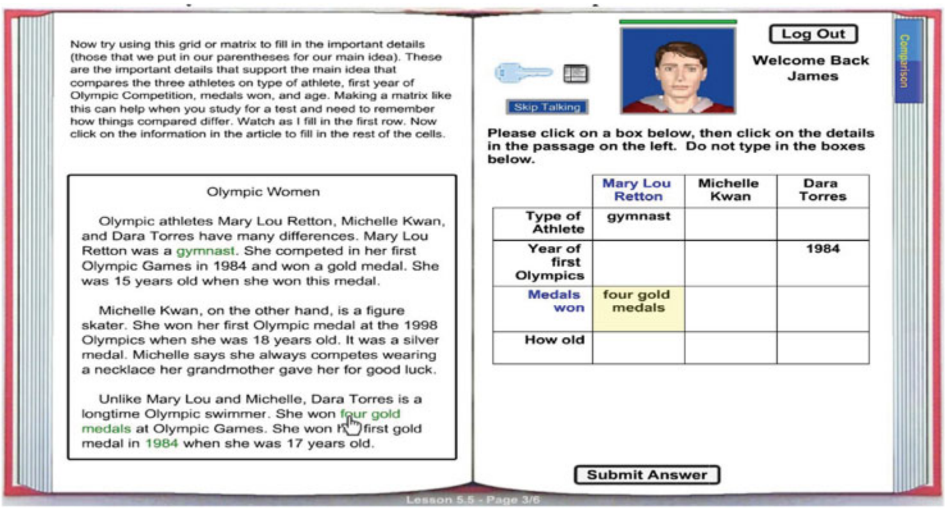

Figure 2. Intelligent tutoring system for the structure strategy matrix style main idea presentation for the comparison text structure.

As the lesson proceeds to the second step of the structure strategy, students are asked to write a main idea. In some lessons, students fill out a matrix (Figure 2) showing how a main idea in the comparison text structure is organized. In the final step, students write a full recall of the passage after the passage has been removed from the screen. The copy and paste functions of the browser and text are disabled so that students must construct their own answers. In the main idea and full recall questions, I.T. checks whether the student is gaming the system by checking for nonsense answers, blanks, and even inappropriate language. I.T. also varies the number of tries the student gets for each answer so that he/she does not always expect the correct answer to be shown in a pop-up window after the third try.

The following steps are built into ITSS using different formats:

1. Identify the overall top-level structure of expository text (e.g., Comparison, Problem and Solution, Cause and Effect, Sequence, and Description) by identifying signaling words (Meyer, 1975) used in text to explicitly cue these structures (such as “in contrast,” “on the other hand,” and “different” for the comparison structure).
2. Write the main idea using patterns for each of the different text structures. For example, for the comparison structure the pattern is: ______ and ______ (two or more ideas) were compared on ______, ________, and ______ (X number of issues compared).
3. Organize reading comprehension and recall by using the structure and main idea (Figure 1).

ITSS has approximately 12 lessons for each text structure beginning with a short passage on favorite presidents of the United States—comparing Abraham Lincoln and George Washington (Social Studies). This lesson is followed by lessons in science (e.g., Differences between African and Indian Elephants) and articles on sports (e.g., Comparing Olympic Athletes). As students complete the comparison text structure followed by the problem/solution structure, they have lessons that combine both text structures. Varying the domains for the reading passages shows learners how text structure can be applied to any domain and helps them transfer their learning (Theodorou, 2006). Combining text structures allows the learner to see how many real-world texts are created.

Motivation is an important aspect of ITSS. As students complete the comparison lessons, they receive a trophy on the screen for their successful completion of five lessons. After the first 12 lessons they get to play a game built into the system. Assessment has an important influence on the strategies, motivation, and learning outcomes of students (Crooks, 1988). ITSS has a complex scheme for the assessment of student responses to full recall and main idea questions. As each student types in his or her full responses, the system parses the answer, classifies the words, checks spelling, and checks synonyms. After this data-cleaning process is completed, the words are matched against a parse tree for the passage. During this phase, the words are classified into the focus areas in the structure strategy (signaling words, main ideas, top-level-structure, and details). Scores are generated for each area, and appropriate feedback is selected and delivered to the learner.

As student responses are collected into the corpus of texts, the system creates new pathways for transitioning students from each interaction. ITSS focuses on pre-defined text passages and responses within a limited corpus of text.

Extension of Structure Strategy Research With ITSS

The current investigation extended previous research in the use of structure strategy instruction. First, it extended the work to larger numbers of students within the research setting of a multisite cluster randomized controlled trial. Second, it extended the work to be delivered via a web-based intelligent tutoring system with a focus on structured, explicit, and scaffolded instruction and feedback adapted to the learner.

Third, this study is the first large-scale efficacy study of the structure strategy instruction provided to children in fifth grade. Fourth, it extended the work to both rural and suburban schools across two states. Fifth, it compared ITSS to standard practice in reading comprehension instruction in many schools.

Based on the theory and supporting research studies we hypothesized that *children learning to read and comprehend expository texts using the structure strategy will show significant improvements compared to students who use other strategies*, such as developing background knowledge and summarizing used in most language arts curricula (business as usual control group). We also explore some factors that may affect the intervention's outcomes variably, such as gender (Halpern, 2006), prior knowledge (van den Broek, Rapp, Kendeou, 2005), and efficacy of usage (Connor et al., 2011).

RESEARCH QUESTIONS

The current study answered the following primary research question. Do students in Grade 5 classrooms using the ITSS delivery of the structure strategy as a partial substitute for the standard language arts curriculum outperform students in control classrooms on standardized and researcher designed measures of reading comprehension?

The study also posed six secondary questions to study whether the effect of ITSS delivered instruction about the structure strategy on reading comprehension varies depending on other factors, such as gender and prior knowledge: Does the effect of ITSS on reading comprehension differ between male and female students? Does the effect of ITSS on reading comprehension differ between low- and medium/high-scoring students on a reading comprehension pretest? Does the effect of ITSS on reading comprehension depend on students' initial reading level? Does the effect of ITSS on reading comprehension vary across rural versus suburban areas? Does the effect of ITSS on reading comprehension vary across schools*?* Do students who used the ITSS system for longer times and with greater frequency perform better on the posttest than students who used it less?

METHOD

Design

This efficacy study investigated the effects of a web-based tutoring system to teach the structure strategy to fifth-grade students in rural and suburban settings. To accomplish this, we used a multisite cluster randomized trial. A volunteer sample of 128 teachers and their classrooms were randomly assigned to the ITSS and control groups within schools. The within-school random assignment of classrooms maintains curricular consistency between the intervention and control classrooms in the same school. Teachers in the ITSS condition agreed to use ITSS for about 20 to 30 min a week instead of regular classroom instruction. The control teachers used the school's standard language arts curriculum for the total language arts instructional time. Schools signed a memorandum of understanding agreeing to participate in the research study.

The within-school random assignment of classrooms to use ITSS or be in the control group required fewer participating schools (based on the power analysis). In addition, this design allowed the research team to easily recruit schools because all teachers at the grade level could receive professional development at the end of the study and would have teacher peers who had already used the software in the schools. One potential disadvantage to this design was the possibility of contamination between intervention and control classrooms. The ITSS software was only accessible using a username and a password, thereby preventing any unauthorized access of the system. Therefore the contamination possibility was minimized.

The statistical power analysis indicated that a minimum of 52 teachers/classrooms were required. To provide a buffer against potential attrition-related problems, the study planned to recruit 58 classrooms (10% more than required) and 1,160 students (assumed average of 20 students per classroom) to detect a 0.2 standard deviation difference between intervention and control classrooms on postintervention reading comprehension. Our recruitment effort resulted in 58 rural and 70 suburban classrooms.

Participants

Recruitment was done initially through letters to all elementary schools in the Common Core of Data–Pennsylvania (U.S. Department of Education, 2008). The PI and Laboratory Extension Specialists followed up with phone calls and presentations to more than 100 schools and intermediate units. These activities resulted in 60 informal expressions of interest from districts.

The main requirement for schools to participate in the research study was the availability of computers to support a one-to-one student–computer ratio. All schools in

Table 2. Mean characteristics of the 45 participating schools

Characteristics	Sample *M*
Suburban schools[a]	
Proportion of racial/ethnic minority students (%)	8
Proportion of socioeconomically disadvantaged students (%)[b]	39
Student education expenditure rate (dollars)[c]	12,037
Rural schools[a]	
Proportion of racial/ethnic minority students (%)	14
Proportion of socioeconomically disadvantaged students (%)[b]	44
Student education expenditure rate (dollars)[c]	12,145

Note. The number of total reporting schools in each state was used as the weight.
[a]Data were obtained from School Data Direct (http://www.schooldatadirect.org) on January 14, 2009. [b]Defined as students eligible for free or reduced-priced lunch. [c]Defined broadly as expenditures per student for the academic component of his/her schooling (excluding costs like transportation). An example of the calculation of this rate is available at http://www.pde.state.pa.us/school_acct/cwp/view.asp?a=182&q=54624

Pennsylvania were eligible because the State had created a high-speed network and had instituted a one to one student computer ratio in 2008. The second requirement was the availability of sufficient bandwidth to support the lesson material delivery through the Internet connections at the school. Again, the schools in Pennsylvania had access to the high-speed network. Some local school networks and firewalls limited the bandwidth and the research team had to follow-up with site visits to confirm the available bandwidth and finalize implementation.

At the conclusion of the recruitment effort a total of 45 schools (22 rural and 23 suburban schools with a total of 128 classrooms) signed the agreement to participate in the research study. Table 2 presents a summary of demographic statistics about the participating schools. Participating schools had an average of 15 students per teacher in both rural and suburban settings. As shown in Table 2, the suburban schools' student population was 14% racial/ethnic minorities and the rural schools' student population was 8% racial/ethnic minorities.

Incentives to participate in the study included the free use of the ITSS software for the study year as well as a 2nd year. In addition, teacher aides were provided to schools to assist in the setup of the computer labs for student use.

All 128 fifth-grade teachers in the participating schools were invited to participate, and none declined. Within school random assignment of 128 classrooms (teachers) to ITSS and control groups was conducted by the research team's methodologist. For example, if a school had four fifth-grade classrooms, two classrooms were assigned to use ITSS and the other two were assigned to the control group. Schools that had an odd number of classrooms or single classrooms were matched with similar schools within the same school district to form a block, and classrooms within a block were randomly assigned to treatment conditions. The random assignment was conducted after the schools had completed their student classroom assignment so that there was no chance of students being placed into the ITSS or control classroom as a result of the research project. Students were automatically assigned to the research condition to which the teacher/classroom was assigned. All students in the fifth grade of participating schools were invited to participate. Each school mailed parental consent forms to all students at the fifth-grade level prior to notification of random assignment. Consent was obtained for 98% of the students invited to participate. At pretest 65 classrooms with 1,351 students were assigned to the ITSS research condition and 63 classrooms with 1,294 students were assigned to be in the control condition.

Procedure

Measures of reading comprehension (standardized reading comprehension test and researcher designed measures) and questionnaires (motivation, self-efficacy) were administered (to both ITSS and control groups) during the pretest before training began in October 2009.

Students in the intervention condition used ITSS for one or two sessions a week for 30 to 45 min each week over a 6- to 7-month period (October 2009 to April 2010). At the beginning of every session, each student picked up his or her ITSS folder containing any instructions, username, password, and earphones and sat individually at the computer. The student opened a browser and logged in using his or her individual username and password. The ITSS software retrieved the student's last login information and saved lesson and page numbers, and then placed the student in the next lesson or page. The student interacted with the ITSS program at his or her own pace, listening to I.T., writing answers, getting feedback from I.T., and getting help when necessary from pop-up windows. At the conclusion of the class period the student logged out and the system saved his or her lesson and page number for use at the next session.

To ensure fidelity of the treatment, the researchers and trained coordinators conducted two formal observations in both the ITSS and control classrooms. The main focus of these observations was to ensure that there was no contamination to the control classrooms and to identify the language arts curriculum being used in the classrooms. Observations were conducted during ITSS use in computer labs and during regular classroom instruction and documented the language arts curriculum used by the schools. The observations also confirmed that within each school the same language arts curriculum was used by the ITSS and control classrooms.

The research team generated biweekly reports for the teachers showing student progress in ITSS. The reports listed one line for each student in the teacher's classroom and showed the last lesson completed, scores on most recent assessments, and the student responses. Gaming was defined as nonsense answers, repeatedly submitting the same answer, and/or submitting blank answers. This allowed the teachers to monitor the students' progress and intervene where students appeared to be gaming the system rather than learning. If a student was provided multiple trials for each lesson then the teacher was consulted to confirm that there were no other issues (e.g., problems with the headphones, disabilities that needed to be addressed by a classroom aide).

To further ensure fidelity of implementation, the research team reviewed computer log files weekly. The computer logs were sorted by time on task for each child and classrooms. If the time on task showed less than 30 min each week the research coordinators contacted the teachers to inquire about the usage. If there were any technology or software challenges, a member of the research team followed up via phone or visited the school to resolve the issues.

Posttest measures (reading comprehension and questionnaires) were administered at the end of the school year April to May 2010 under the same conditions as pretest administration.

Table 3. Grade 5 measures and reliability estimates

	Reliability Estimates	
Measures	Pretest	Posttest
Reading comprehension		
Gray Silent Reading Test[a]	.88	.88
Problem and solution text[b]		
Total recall	—	98%
Competency	—	89%
Comparison text[b]		
Total recall	—	99.1%
Competency	88%	89%
Main idea quality	92%	93%
Signaling test[c]		98%
Affective measures[a]		
Computer attitudes	.81	.72
Learning self-efficacy	.62	.62
Reading self-efficacy	.77	.78
Structure strategy self-efficacy	.75	.75

[a]Cronbach's alpha. [b]Interrater agreement. [c]The pretest represents Computer Scored Signaling Tests.

Materials

The outcome measures and their estimated reliability indices are listed in Table 3. Reliability estimates were calculated based on data collected for this study.

Cognitive Outcome for Research Questions

Reading comprehension was measured using a standardized reading comprehension test with multiple-choice questions primarily about short narrative texts. Reading comprehension also was measured using experimenter-designed recall and main idea tests about expository texts.

Standardized Test of Reading Comprehension. The Gray Silent Reading Test (GSRT; Wiederholt & Blalock, 2000) Form B was administered at pretest and Form A was administered at posttest. Pretest score on the GSRT was used as a covariate for data analyses used to examine the effects of ITSS instruction on our dependent measures that focus on reading comprehension. Cronbach's alpha for both forms of the GSRT was reasonably high ($\alpha =$ .88).

The GSRT was selected because (a) group administration due to budget, time, and school schedule constraints; (b) testing of deep comprehension processes with recall, application, and inference questions; (c) use of the same test for various grade levels; and (d) at least two forms with good psychometric properties. In the manual reliability data for the alternate forms are provided at 12 age intervals for the two forms given one following the other; the average alternate form reliability was .85 (.87 for 10-year-olds) and delayed alternate-form reliability was reported at .83. Coefficient alpha reported for forms A and B were .95 and .94, respectively.

Experimenter-Designed Measures of Reading Comprehension. Two equivalent test forms were created (Meyer et al., 2010) and one was administered before the children started ITSS and the second immediately after completing the program to test students' understanding of expository texts with problem and solution and comparison text structures. The problem and solution set of two equivalent passages had 98 words, 72 idea units, and equivalent scores on traditional measures of readability, text structure, and signaling (see Meyer, 2003). Each text presented a relatively unfamiliar problem and its cause and a solution that eliminated the cause of the problem on topics of rats or dogs. The article about rats was an authentic newspaper article (see Meyer & Poon, 2001). Students were asked to recall all they can remember after reading each problem and solution text and placing it out of sight in an envelope. Table 3 shows the interrater agreement between two scorers for this free recall task with the problem and solution set of texts (89–98%).

A set of two passages were also prepared for the comparison structure: (a) Pygmy versus Emperor Monkeys and (b) Adelie versus Emperor Penguins. Each comparison passage had 128 words, 15 sentences, and 96 idea units. As seen in Table 3 there are two tasks for the comparison structure: (a) a recall task similar to the one used for the problem and solution set of articles and (b) a comparison main idea task where the student is asked to write a two-sentence main idea with the text available for consultation. Table 3 also shows high interrater agreement for the measures collected for the comparison free recall and main idea tasks (88–99%).

Competency ratings for use of the problem and solution and comparison structures (scores of 1–8) were assessed to determine whether a student was proficiently using a text structure as outlined in the ITSS program. Competence ratings assessed use of the structure to organize correct ideas (see Meyer et al., 2010, for first use of the competence measure). Maximum competence in using problem and solution involved a recall that specified the correct problem and its cause as well as the solution posited in the original text proposed to prevent the cause (8 on a 8-point scale). A score of 1, representing no competence using the problem and solution structure, indicated that the student's recall included no problem, no solution, and no cause related to the text (e.g., "researchers work with rats and mice"). A score of 2 was given for a recall containing a signaled cause but no problem and no solution. Content information about the problem may have been mentioned, but the problem was not developed (e.g., "urine causes allergies"). A score of 3 indicated a student's recall accurately identified just one part of the problem and solution structure (usually just the problem and rarely just the solution). A score of 4 was assigned to a recall that contained the problem and cause but no solution. Alternatively, a score of 4 was assigned to a recall with correct problem(s) with its correct cause(s), but the solution was incorrect. Signaling was not required for a correct problem score. A score of 5 indicated that a student had a problem part in his or her recall and a solution part with the correct content of the problem (e.g., "researchers are getting sick from working with rats") and correct content of the solution (e.g., "kindness to rats"). A score of 7 indicated that a student organized his or her recall like that of a score of 5 but additionally presented the cause of problem when discussing the problem (e.g., "allergies caused by rat's urine"). A score of 6 suggested the cause of the problem only in the solution part of the recall (e.g., "kindness to rats makes them not pee"). The maximum score of 8 included the requirements of scores 6 and 7 (e.g., "Scientists who work with rats and mice often develop allergic reactions to them. This is a danger to the researchers. It was discovered that the reaction was to a protein in the urine. A meeting by the Organization of Health met to discuss this and Dr. Slovak suggested that if the researchers were nice to the rats and mice, they would be less likely to be splattered with the rat's urine.").

In a similar manner, competence in using the comparison structure was accessed on both the free recall task (comparison competency on 8-point scale) and two-sentence main idea task with the text available to consult (comparison main idea quality on 6-point scale).

Affective Outcomes for Research Questions

Due to constraints in amount of time available for testing in the various school districts of this study, we needed to use a subset of affective measures that we had used in past research (e.g., Meyer et al., 2002, 2010). We conducted factor analyses from our past studies on complete scales to identify four to six items from each scale that best indicated affective variables of interest with Cronbach alpha (Cronbach, 1951) of at least .70. At pre- and posttests, students completed six items related to computer use and opinion referred to as computer attitude (Krauss & Hoyer, 1984). They also completed four items from the Self-concept as a Reader subscale of Motivation to Read profile referred to as reading self-concept (Items 7, 9, 13, and 15 from Gambrell, Palmer, Codling, & Mazzoni, 1996). In addition, they completed five items about learning self-efficacy (Items 9, 12, 13, 18, and 20 from Sherer et al., 1982) and five experimenter-designed items related to the student's self-efficacy for using signaling words and writing good main ideas (structure strategy self-efficacy). Scores from all questionnaire items were subjected to an exploratory factor analysis, separately for pretest and posttest. Cronbach's alphas (Cronbach, 1951) for the questionnaires measuring reading self-concept, learning self-efficacy, structure strategy self-efficacy, and computer attitudes are shown at the bottom of Table 3. For both pretests and posttests, all items consistently loaded on their respective factors except for one from the learning self-efficacy factor (When I decide to do something, I go right to work on it) and one from the computer use factor (Do you prefer learning on a computer than learning in the classroom). These items were therefore dropped from their respective subscales. Cronbach's alphas for the resulting subscales ranged from .72 to .78 on the pretest and posttests (.72–.81 for computer attitude, .77–.78 for reading self-concept, and .75 for structure strategy self-efficacy). The Cronbach's alpha for the measure of learning self-efficacy was .62, which falls in the lower limit of acceptability between .60 and .70 (Cronbach, 1951).

Scoring

The graduate students in educational psychology who scored the main idea and recall tasks were blind to the experimental condition of the participants. The graduate students were trained in Meyer's (1975, 1985) scoring procedure based on a propositional analysis of the ideas in text with interrelationships among ideas specified in a hierarchical content structure. Scoring manuals based on Meyer's approach to discourse analysis were prepared for each passage and task. Scoring structures were typed into an adapted Microsoft Excel program to score and automatically tally idea units from the texts and the interrelationships among these idea units. Tallies for each protocol were entered into the dataset for total recall. The graduate students received extensive, mentored training in the scoring procedures for all experimenter-designed tasks by an experienced researcher/professor in educational psychology. At least 10% of the data from each of the measures were randomly selected from the conditions and times of testing to check interrater agreements.

Competency ratings for use of the problem-and-solution and comparison structures (scores = 1–8) were assessed to determine the degree to which a fifth-grade student proficiently used the text structure as outlined in the ITSS program. The competency ratings assessed use of the structure to organize correct content. Four graduate students scored the comparison main idea quality for both the pretest and posttest. Agreement among all four scorers on these measures was 92% on the pretest and 93.3% for the posttest main idea task. Then these four graduate students were grouped in teams of two to score the recall data from the comparison text or the problem-and-solution text.

A team of two graduate students scored each of the recalls from the comparison recall text for the first one fourth of the data. Their scores were compared, and the inconsistencies between the two scorers were resolved by discussion under the guidance of an experienced researcher in educational psychology. Ten percent of this one fourth of the data was randomly selected to calculate the percentage agreement. The rest of the data was divided into two parts with equal amount (odd and even numbers from a numerical list of data). Each scorer scored one part of the data and randomly selected 10% of the other part of data to score. The 10% of the first one fourth of the data and 10% of the rest of the data were combined to calculate overall percentage agreements between the scorers for all posttest recall measures. The percentage of agreement between scorers for comparison competency was 88.9%. Percentage of agreement between scorers for total recall was 99%.

Similarly, a set of two measures were obtained from the problem-and-solution recall posttest, including competency ratings and total recall. The percentage for agreements between scorers for problem-and-solution competency rating was 89%. Percentage agreement between scorers for total recall was 98%.

The signaling posttest was scored on a 7-point scale for each of four missing signaling words in the comparison main idea task only. Scores per missing signaling word ranged from verbatim use of the intended signaling word (7 points, e.g., different) to a content word (1 point, e.g., bake) that made little semantic or grammatical sense. A trained graduate student scored the signaling test; a stratified random sample of 10% of the data from the signaling test was scored by another trained graduate student. Percentage agreement was 98%.

Data Analysis

To determine if there were differences between ITSS and control classrooms with respect to reading performance, a series of hierarchical linear modeling (HLM; Raudenbush & Bryk, 2002) equations were specified. A set of analyses was run for each of the primary dependent variables (GSRT and researcher-designed measures of reading comprehension) using the HLM7 software program. Missing data were deleted listwise at the time of analysis for each model to maximize the use of available data.

HLM Model Specifications: Addressing Research Questions

For the HLM models, students are nested in classrooms and classrooms are nested within schools. An unconditional three-level model (M0) was first estimated to gauge the variability due to each level. A base model was then estimated to answer the main research question, in which there were predictor variables at each level. At the student level, predictor variables included gender (1 = female, 0 = male; grand-mean-centered) and reading comprehension covariates. Reading comprehension covariates included group-mean-centered pretest scores on GSRT for all outcome measures and the pretest scores for comparison competency with the outcomes measures for GSRT, problem and solution total recall, problem and solution competency, and comparison competency. For the Signaling test and main idea quality posttests, we used the same corresponding measure from the pretest instead of comparison competency. Because of the large volume of data collected, these covariates were the only researcher designed reading measures scored manually for the pretest. At the classroom level, treatment efficacy was tested using contrast codes for experimental conditions (i.e., ITSS vs. control with codes of $\frac{1}{2}$ and $-\frac{1}{2}$, respectively; these contrast codes were used such that unstandardized regression coefficient corresponded to the difference between the unweighted means of the groups involved in the contrast). Grand-mean-centered classroom-level pretest scores on GSRT and the corresponding researcher-designed pretest measures were included as covariates. At the school level, differences between rural and suburban schools were examined (1 = rural, 0 = suburban; grand-mean-centered). In addition, variance associated with each of the three levels was estimated. This three-level base model (M1) was used to address the main research question of whether ITSS classrooms outperformed control classrooms on reading comprehension after controlling for other relevant factors such as prior reading level, gender, and school locale.

To test whether the effect of ITSS on reading comprehension differed between male and female students, a cross-level interaction between treatment and gender was added to the base model by specifying the level-1 coefficient for gender as a function of treatment (M2). Similarly, a cross-level interaction between treatment and reading pretest level on GSRT (1 = students' grade equivalent scores fell below Grade 5, 0 = students' grade equivalent scores were at or above Grade 5) was added to determine whether the effect of ITSS differs between below grade-level and higher scoring students on a reading comprehension pretest. Note that the student level pretest on the GSRT was replaced by its dichotomized counterpart in this model (M3). To make use of all information afforded by the reading pretest measures (without categorizing scores), cross-level interactions between treatment and each of the reading pretests (GSRT and corresponding researcher-designed pretest measure) were added to the base model to examine a similar question of whether the effect of ITSS on reading comprehension depend on students' initial reading level (M4). To address the question of whether the effect of ITSS on reading comprehension varies across rural/suburban areas, a cross-level interaction between treatment and school locale was added to the base model (M5). Should any of the additional cross-level interaction terms be statistically significant, we would further examine the pattern of the interaction. Otherwise, the more parsimonious base model would be interpreted.

To test whether ITSS had different effects in different schools rather than having a common effect across all schools, we estimated variability of treatment effect across schools by modeling the level-2 coefficients for treatment as random effects. Should their variance estimates be statistically significant, we would estimate the 95% plausible value range of treatment effect among schools. On the other hand, if the variance estimates were nonsignificant, we would retain the fixed effects model for parsimony.

In addition, we estimated effect sizes of ITSS as compared to the control based on the base model (M1). Specifically, we computed the effect size as a standardized mean difference by dividing the adjusted (for pretest scores and other covariates) group mean difference by the unadjusted *pooled* within-group student level standard deviation of the outcome measure or where applicable, by the pooled within-group student level standard deviation of the pretest scores.

To examine the effects of ITSS on students' affective measures we used a similar three-level structure as Model 1. The affective measures studied included the reading self-concept, learning self-efficacy, structure strategy self-efficacy, and computer attitudes. Covariates

Table 4. Student- and class-level means and standard deviations on the Gray Silent Reading Test

	Treatment Condition					
	ITSS			Control		
	n	*M*	*SD*	*n*	*M*	*SD*
Class level						
Pretest	65	27.82	4.07	63	27.06	4.18
Posttest	65	35.49	3.37	62	32.64	3.84
Student level						
Pretest	1,351	28.04	11.45	1,294	27.29	11.44
Posttest	1,258	35.43	11.13	1,227	32.84	12.21

Note. ITSS = intelligent web-based tutoring system for the structure strategy.

included all four affective pretest measures and the GSRT (both student and classroom levels) as well as gender and school locale.

Last, we examined simple Pearson correlations between the GSRT posttest and each of the indicators of system usage (average minutes used per week and total number of ITSS questions answered). A significant positive correlation would indicate that students who used the system more performed better on posttest.

RESULTS

There was no statistically significant difference between ITSS and control groups on the pretests at the random assignment classroom level, $t(126) = |1.05|$, $p = .29$ for the GSRT; $t(126) = |.06|$, $p = .96$ for the Signaling test; $t(110) = |.97|$, $p = .33$ for Main Idea Quality; $t(122) = |1.08|$, $p = .28$, for Comparison Competency. This indicated that the ITSS and control classrooms were comparable in their reading level before the implementation of the experiment.

Class- and student-level simple descriptive statistics by treatment condition for GSRT and experimenter-designed reading comprehension measures are presented in Tables 4 and 5, respectively. Statistical test results of treatment effect from HLM analyses and effect sizes on GSRT, comparison, and problem and solution posttest scores are summarized in Table 6. HLM analyses (M0-M6) were conducted on each of the reading comprehension measures. However, for concern of space, we only present complete HLM results on the GSRT post-test (see Table 7). Effect estimates for ITSS presented in Table 6 were extracted from M1 for each of the outcome measures. Results are discussed by research questions.

Primary Research Question

To address the question of whether Grade 5 classrooms using the ITSS delivery of the structure strategy as a partial substitute for the standard language arts curriculum outperformed control classrooms on standardized and researcher designed measures of reading comprehension, we used results from HLM Model 1. Students in ITSS classrooms on average scored 2.34 points (or .2 standard deviations) higher on GSRT adjusted posttest scores and

Table 5. Class- and student-level means and standard deviations on experimenter-designed measures

	Treatment Condition					
	ITSS			Control		
Measure	*n*	*M*	*SD*	*n*	*M*	*SD*
Class level						
Comparison text						
Signaling test	64	16.44	5.76	63	13.28	5.33
Main idea quality	56	3.63	.45	53	2.84	.61
Total recall	56	31.12	7.39	53	25.02	6.57
Comparison competence	56	4.72	.84	53	4.05	.91
Problem and solution text						
Total recall	56	20.81	4.19	53	17.85	4.42
Problem and solution competence	56	4.06	.73	53	3.61	.94
Student level						
Comparison text						
Signaling test	1,365	16.48	8.64	1,300	13.38	8.47
Main idea quality	1,156	3.61	1.58	1,004	2.84	1.50
Total recall	1,161	31.31	18.34	1,012	24.87	16.48
Comparison competence	1,159	4.73	2.44	1,004	4.01	2.41
Problem and solution text						
Total recall	1,163	20.82	12.66	1,017	17.78	11.66
Problem and solution competence	1,163	4.07	2.52	1,017	3.58	2.38

Note. ITSS = intelligent web-based tutoring system for the structure strategy.

2.93 points (or .42 standard deviations) higher on comparison Signaling posttest scores (see Table 6) than students in control classrooms holding reading pretest scores, gender, and school locale constant. These differences were statistically significant at $p < .05$. Adjusted posttest scores were also statistically significantly higher for students in ITSS classrooms than their control counterparts on all other researcher-designed reading comprehension measures (see Table 6): problem and solution competency (adjusted difference = .37, ES = .15), problem and solution total recall (adjusted difference = 2.38, ES = .2), comparison main idea quality (adjusted difference = .82, ES = .53), comparison total recall (adjusted difference = 5.57, ES = .32), and comparison competence (adjusted difference = .64, ES = .26). The effect size of .20 on the standardized GSRT test was considered small, whereas the effect size of .53 on the comparison main idea quality was considered medium.

Secondary Question 1

Results from Models M3 and M4, respectively, provided an answer to the research question on whether the effect of ITSS on reading comprehension differed between low- and

Table 6. Effect sizes of ITSS on reading measures

Measures	Coefficient for ITSS (*SE*) from HLM [a]	Pooled Student-Level *SD*	Effect Size
Gray Silent Reading Test	2.34*** (.44)	11.45	.20
Comparison text			
Signaling test	2.93*** (.39)	7.05	.42
Main idea quality	.82*** (.08)	1.54	.53
Total recall	5.57*** (1.07)	17.49	.32
Comparison competence	.64*** (.14)	2.42	.26
Problem and solution text			
Total recall	2.38*** (.64)	12.20	.20
Problem and solution competence	.37** (.12)	2.45	.15

Note. Effect size = Adjusted difference between ITSS (coded $\frac{1}{2}$) and control (coded $-\frac{1}{2}$) groups divided by the student-level pooled standard deviation. ITSS = intelligent web-based tutoring system for the structure strategy; HLM = hierarchical linear modeling.
[a]Estimates are extracted from model 1; degrees of freedom = 77.
**$p < .01$. ***$p < .001$.

medium/high-scoring students on the reading comprehension pretest and whether the effect of ITSS on reading comprehension depended on students' initial reading level.

For the GSRT, the interaction between the student-level pretest and ITSS was significant at the .05 level regardless of whether the pretest was categorized as dichotomous (Model M3 in Table 7) or continuous (Model M4 in Table 7). This indicated that the effect of ITSS, adjusted for other covariates in the model, varied depending on students' initial reading level as shown in Figure 3. Students who performed below grade level at pretest made larger gains in the posttest than did students who performed initially at or above grade level. There were no statistically significant interaction effects on the experimenter-designed reading measures.

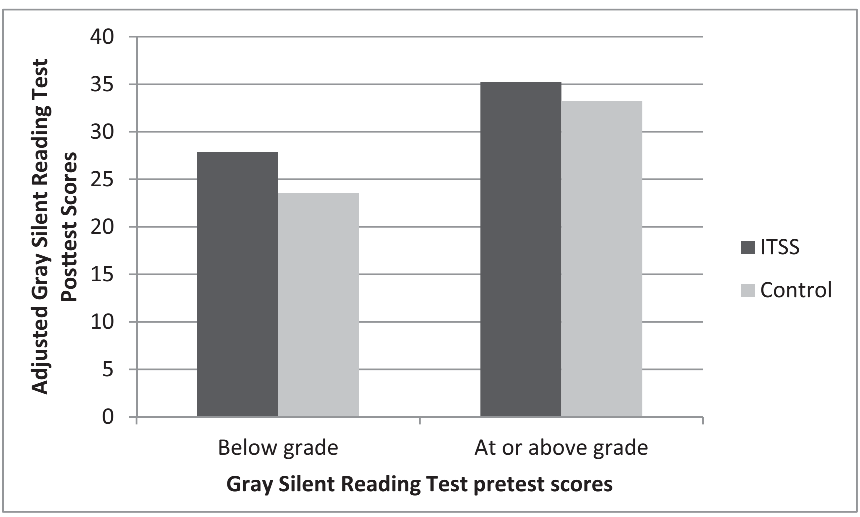

Figure 3. Cross-level interaction between ITSS and student pretest level on adjusted gray silent reading test posttest scores.

Table 7. HLM results on Gray Silent Reading Test (GSRT) posttest scores for Grade 5

	M0	M1	M2	M3	M4	M5	M6
Fixed effects estimates (SE, *df*)							
Intercept	34.10*** (.38, 31)	34.34*** (.38, 30)	34.34*** (.38, 30)	34.23*** (.35, 30)	34.34*** (.38, 30)	34.34*** (.38, 30)	34.32*** (.38, 30)
Female	—	–.17 (.38, 2537)	–.17 (.38, 2536)	.27 (.37, 2536)	–.15 (.38, 2535)	–.17 (.38, 2537)	–.18 (.37, 2505)
GSRT pretest[a]	—	.51*** (.03, 2537)	.51*** (.03, 2536)	–8.15*** (.51, 2536)	.50*** (.03, 2535)	.51*** (.03, 2537)	.51*** (.03, 2505)
Comparison competency on pretest	—	.90*** (.12, 2537)	.90*** (.12, 2536)	1.24*** (.11, 2536)	.89*** (.12, 2535)	.90*** (.12, 2537)	.90*** (.12, 2505)
Class average GSRT pretest	—	.63*** (.07, 77)	.63*** (.07, 77)	.38*** (.08, 77)	.63*** (.07, 77)	.63*** (.07, 76)	.64*** (.07, 46)
Class average comparison competency on pretest	—	–.05 (.30, 77)	–.05 (.30, 77)	.04 (.33, 77)	–.05 (.30, 77)	–.05 (.30, 76)	–.05 (.32, 46)
Rural	—	.30 (.78, 30)	.30 (.78, 30)	.32 (.72, 30)	.30 (.78, 30)	.29 (.78, 30)	.26 (.73, 30)
ITSS	—	2.34*** (.44, 77)	2.34*** (.45, 77)	2.02*** (.42, 77)	2.34*** (.44, 77)	2.34*** (.45, 76)	2.28*** (.44, 31)
ITSS × female	—	—	.12 (.75, 2536)	—		—	—
ITSS × GSRT pretest[a]	—	—	—	2.32* (.94, 2536)	–.11* (.04, 2535)	—	—
ITSS × comparison competency on pretest	—	—	—	—	.22 (.27, 2535)	—	—
ITSS x rural	—	—	—	—	—	–.08 (.93, 76)	—
Variances for random effects (*df*)							
Sites	.64 (31)	2.46*** (30)	2.46*** (30)	1.83*** (30)	2.46*** (30)	2.45*** (30)	2.57*** (29)
Classrooms	6.84*** (79)	.80 (76)	.80 (76)	.04 (76)	.81 (76)	.80 (76)	.55 (46)
Students	130.61	90.34	90.33	102.18	90.02	90.34	90.34
ITSS	—	—	—	—	—	—	.57 (30)
Model fit statistics							
Deviance	16882.89	14476.04	14476.02	14698.39	14469.51	14476.03	14473.99
No. of parameters	4	11	12	12	13	12	13

Note. M = model; ITSS = treatment condition ($\frac{1}{2}$ = ITSS, $-\frac{1}{2}$ = control); female = gender indicator (1 = female, 0 = male, grand-mean-centered); rural = locale indicator (1 = rural, 0 = suburban, grand-mean-centered).

[a]For M3, GSRT pretest scores are classified into reading below grade level or not.

$^{*}p < .05$. $^{**}p < .01$. $^{***}p < .001$.

Secondary Question 2

The HLM model M6 addressed the question of whether the effect of ITSS on reading comprehension varied across schools. The estimated variance of adjusted ITSS effects across schools on the GSRT posttest and all experimenter-designed reading measures was not statistically significantly different from zero at the .05 level. Difference in deviance between the random ITSS effect model (M6) and the fixed ITSS effect model (M1) on these measures was also not statistically significant. In other words, there was not sufficient evidence to suggest that adjusted ITSS effects (for the covariates) on the GSRT standardized test and researcher-designed measures differed significantly across schools. Therefore, the more parsimonious fixed-effects model was preferred.

Model M1 was the preferred model for researcher-designed tests. It reduced 38% of the classroom level variance in the Signaling test, 55% in the problem and solution total recall posttest, 61% in problem and solution competence, 85% in comparison main idea quality, 54% in comparison total recall, and 74% in comparison competence. The model also reduced 22% of the student level variance in the Signaling, 29% in the problem and solution total recall posttest, 26% in problem solution competence, 12% in comparison main idea quality, 27% in comparison total recall, and 20% in comparison competence.

Secondary Questions 3 to 5

Models M2 and M5, respectively, addressed the research questions on whether the effect of ITSS on reading comprehension differed between male and female students and whether it varied across rural versus suburban areas. The effect of ITSS did not appear to vary as a function of gender or school locale on any of the reading outcomes that we examined. Moreover, neither students' gender nor schools' locale seemed to make a significant difference on these posttest reading scores after pretest scores were controlled for (see results from M1).

Secondary Question 6

Finally, Pearson correlations of GSRT reading posttest scores with fidelity measures were calculated to address the question of whether students who used the ITSS system for more time and who answered more questions performed better on the posttest. Average minutes used per week was not significantly related to GSRT posttest scores. However, the total number of questions answered demonstrated a positive and statistically significant correlation with GSRT ($r = .19$). These correlational analysis results suggested that sheer time usage may not be a good indicator of fidelity. Students using extra time could be gaming the system rather than working on the lessons. The actual number of questions answered appeared to be a better indicator of fidelity as it indicated students' effort to learn the lessons.

As a sensitivity analysis, we also reanalyzed the fixed-effect model (M1) by adding affective pretest scores (computer attitude, reading self-concept, learning self-efficacy, and structure strategy self-efficacy) as covariates for the GSRT. Magnitudes of the adjusted ITSS effects on the standardized reading comprehension (GSRT) outcome measure remained about the same. Hence, detailed results of this analysis are not presented to conserve space. However, it might be worth noting that both student- and classroom-level learning self-efficacy and reading self-concept pretest scores were significant predictors ($p < .05$) for GSRT posttest scores holding reading pretest scores constant. In addition, we analyzed the effect of ITSS on each of the affective posttest measures with affective pretest scores and reading pretest scores controlled. The results indicated that ITSS had a significant direct effect on posttest scores for reading self-concept (adjusted difference = .06, SE = .02, $p = .01$; ES = .12) and structure strategy self-efficacy (adjusted difference = .06, SE = .03, $p = .04$; ES = .09), but not on the posttest measures of learning self-efficacy (adjusted difference = –.01, SE = .02, $p = .66$; ES = −.02) and computer attitude (adjusted difference = −.02, SE = .06, $p = .70$; ES = −.02).

In summary, ITSS appeared to have a nontrivial positive influence on reading comprehension outcome measures above and over increases that could be predicted by students' initial reading and affective levels. Moreover, ITSS seemed to have a small but statistically significant positive direct effect on reading self-concept and structure strategy self-efficacy measures.

DISCUSSION

In this study, we extended the investigation of the efficacy of structure strategy instruction to larger numbers of students in a multisite cluster randomized controlled trial, conducted in both rural and suburban settings across two states. This instruction was delivered via ITSS, a web-based intelligent tutoring system with a focus on structured and explicit modeling, practice tasks with scaffolding, and focused feedback based on the learner's performance and attempts. In addition, we tested the effects of ITSS on the largest sample of fifth-grade students in the extant literature in a fully randomized controlled trial. The primary research question was whether fifth-grade classrooms using ITSS delivery of the structure strategy as a partial substitute for standard language arts curriculum outperform control classrooms on standardized and experimenter-designed measures of reading comprehension.

The findings from this large-scale randomized controlled trial show that students in ITSS classrooms statistically significantly outperformed their control classroom counterparts on the standardized reading comprehension test (GSRT) and the researcher-designed measures. The effect size for the GSRT was small and the effect sizes for the researcher-designed measures particularly targeted by the instruction (i.e., main ideas and signaling words) were moderate. It is noteworthy that these results were obtained when logs of students use of the system revealed that overall students actually only used ITSS for one classroom period (30–45 min) a week for approximately 30 weeks. Together with results from previous studies (Meyer et al., 2010; Meyer, Wijekumar, & Lin, 2011; Wijekumar, Meyer, & Lei, 2012) this study extends the accumulating evidence that the structure strategy delivered via ITSS is a likely cause for the improvements in reading comprehension. Additionally, these findings combined with previous methods of delivering the structure strategy (e.g., Meyer & Poon, 2001; Meyer et al., 2002) suggest that the structure strategy can be causally linked to improvements in reading comprehension.

These results should also be viewed in the context of some of the recent large scale randomized controlled trials on reading programs at upper elementary and middle grade levels (James-Burdumy et al., 2009). The U.S. Department of Education commissioned researchers to conduct large-scale randomized controlled trials on four reading products: Project CRISS, ReadAbout (Scholastic), Read for Real (Zaner-Bloser), and Reading for Knowledge (Success for All). Findings from these studies showed no significant differences between the control classrooms and classrooms using any of these products. Drummond et al. (2011) conducted a study on the impact of Thinking Reader software on sixth-grade reading, vocabulary, comprehension, strategies, and motivation. Results showed a .03 effect size on reading comprehension. Slavin, Cheung, Groff, and Lake (2008) reviewed studies in middle and upper-middle schools on reading interventions delivered by teachers as well as computers and found effect sizes ranging from .00 to .10 for computer aided instruction. In contrast, our findings for ITSS effects are statistically and practically significant. Moreover, the effect sizes we found with ITSS were larger than those reported by James-Burdumy et al. (2009) and Drummond et al. We attribute these findings to the systematic use of the structure strategy in this efficacy study in contrast to the instructional approaches used in the James-Burdumy et al. and Drummond et al. studies (e.g., summarizing, question generation, etc.)

Another recent research study conducted by Connor et al. (2011) presented statistically significant effect sizes of .19 for total score and .20 on reading comprehension using an individualized student instruction learning approach for third graders. The results from this study on ITSS have produced similar effect sizes on the GSRT. Connor et al. used a computer tool to individualize student instruction for third graders. Their approach recommended a wide range of instructional activities for students based on their prior knowledge and skills. Some of the approaches used by Connor et al. in the individualizing were similar to those approaches used in this study (e.g., text structure, schema building, predicting, inferencing). The similarities between their approach and the current study include the use of some text structure instruction, metacognitive prompts, and the smaller grained instruction (even though Connor et al. focused on the teacher delivering instruction to small groups and ITSS provides one-on-one instruction on comprehension in expository texts only). Finally, the Connor et al. study focused on third grade and the current study focused on fifth grade. Results from both these studies may be used to create richer individualizing approaches for students (e.g., using the structure strategy style text structure instruction during individualizing to student needs) and the web-based technologies may provide another avenue beyond reliance on teachers to deliver the individualized instruction in small groups.

The findings reported here were based on approximately 30 min of class time spent using ITSS for 7 months in rural and suburban schools. This weekly intervention delivery is practical for schools and is about a third the dose as in the development grant study (Meyer et al., 2010).

The results from the secondary analyses also showed important findings. First, the effects of the ITSS intervention did not vary significantly for rural versus suburban settings or male versus female students. Also, no differences were found on the outcome measures between male and female students as well as no differences between rural and suburban students. However, students reading below grade level on the pretest showed higher gains on the GSRT posttest than their on- or above-grade-level counterparts. This is likely due to a ceiling effect on the standardized reading comprehension test where more advanced students were unable to show large gains. Grade-level and above readers showed similar gains to below-grade-level readers on the experimenter-designed measures where there were no interactions between treatment and pretest reading level. Often maturation or interventions show larger effects for more skilled students than less skilled readers, where the rich get richer as characterized by the Matthew effect (Stanovich, 1986). It is notable that ITSS yielded more benefit on the standardized test of reading comprehension for below-grade-level readers with the greatest needs for improvement in reading comprehension.

The design of this research study used a within-school random assignment of classrooms to the ITSS or control classrooms. One of the drawbacks of this design is the possibility of contamination of the control classrooms. The ITSS software access was carefully monitored and only students and teachers in the ITSS group were given access with usernames and passwords. In addition, we conducted fidelity observations in the ITSS and control classrooms. It was confirmed that none of the students in the control classrooms were using the ITSS software.

The within-school random assignment of classrooms provides an important advantage ensuring curricular consistency between the ITSS and control classrooms within each school. The hierarchical linear model included schools at the highest level to account for school variability.

Our study was designed to examine the effects on reading comprehension of fifth-grade children learning how to write good main ideas via the structure strategy versus traditional fifth-grade reading comprehension instruction that focuses on activities that assume the children know how to write/find good main ideas. Based on theory and research, we hypothesized that children learning to read and comprehend expository texts using the structure strategy would show significant improvements compared to students who use other strategies, such as highlighting important ideas and question generation, which are used in most language arts curriculum. The findings of our study showed statistically significant effects on the researcher-designed measures that included main idea construction.

In terms of practical significance, effects sizes were greatest for main idea quality and identifying signaling words that cue text structure. Thus, our hypothesis was supported for instruction about text structure via ITSS providing greater boosts for writing good main ideas than the control classrooms. The substantial effect size (i.e., .53) for main idea quality resulted despite an indication that some schools included aspects of the structure strategy in the language arts curriculum. Fidelity checks noted that three schools with 12 classrooms were actively pursuing a text-structure-based curriculum for all their classrooms and had numerous posters and wall displays that showed signaling words and text structure information. Despite this unexpected overlap in some of the content between ITSS and the language arts programs in 9% of the classrooms, ITSS showed statistically significant and moderate effects sizes on measures most directly related to producing good main ideas.

The current findings support our past work indicating that careful work in the ITSS lessons relates to improved reading comprehension. New in this study were the findings that ITSS instruction increased statistically significantly both reading self-concept and structure strategy self-efficacy.

In a recent design feature study with ITSS (Meyer, Wijekumar, & Lin, 2011), a more individually tailored version was developed to provide remediation or enrichment lessons to better match the needs of fifth-grade readers. The individualized version was designed to match the individual learner's needs by using the student's performance during a lesson to adapt the sequence, complexity, and/or text difficulty of the proceeding lesson. The standard ITSS version with individualized feedback, also used in the current study, provided students with the same, fixed sequence of lessons regardless of their performance. Students receiving the more individualized instruction demonstrated higher mastery achievement goals when working in ITSS lessons than students receiving the standard instruction ($d = .53$). Also, fifth-grade students receiving the more individualized ITSS showed greater improvement on the GSRT, the Signaling test, better work in lessons, and more positive posttest attitudes toward computers than students receiving standard instruction. Further individualization of ITSS instruction may yield promising practical results.

The results from this study also support the use of the intelligent web-based delivery of the structure strategy as an effective method to delivering the instruction to a large audience. The ITSS system was compatible with almost all operating systems and hardware systems available in schools. The system also supported more than 6,000 concurrent users without any degradation of performance (because this research was part of a series of studies conducted at multiple grade levels).

These findings can be used to inform reading comprehension instruction in classrooms. The complete set of practice tasks within ITSS provides broad and deep implementations of the structure strategy. For example, the tailored main idea patterns for each text structure and practice within ITSS using texts from science, social studies, and other domains are important features that should be carefully considered when implemented within the classroom.

No one study can address all questions about the efficacy of an intervention. Regardless of rigor, all studies have limitations, especially in terms of generalizability to other settings and contexts. This study focused on fifth-grade classrooms in rural and suburban settings. Therefore the findings relate to the participating schools and classrooms. In addition, the findings apply to the use of ITSS for one class period a week for the academic year.

Future studies are needed to examine the most critical components of ITSS for improving reading comprehension and the use of ITSS as a springboard for teachers to apply the structure strategy with content from their science, math, social studies, and language arts curricula. Other studies should also focus on a more representative sample including urban and high-poverty schools.

REFERENCES

Aleven, V. A., & Koedinger, K. R. (2002). An effective metacognitive strategy: Learning by doing and explaining with a computer-based cognitive tutor. *Cognitive Science*, *26*, 147–179.

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

Anderson, J. R., Corbett, A. T., Koedinger, K. R., & Pelletier, R. (1995). Cognitive tutors: Lessons learned. *The Journal of the Learning Sciences*, *4*, 167–207.

Baylor, A. (2001). The effects of MIMICing instructional theory with MIMIC (multiple intelligent mentors instructing collaboratively), an agent-based learning environment. Retrieved from http://search.proquest.com/docview/62194245?accountid=13158

Biancarosa, G., & Snow, C. (2006). *Reading next—A vision for action and research in middle and high school literacy: A report to Carnegie Corporation of New York* (2nd ed.). Washington, DC: Alliance for Excellent Education.

Cantrell, S. C., Almasi, J. F., Carter, J. C., Rintauma, M., & Madden, A. (2010). The impact of a strategy-based intervention on the comprehension and strategy use of struggling adolescent readers. *Journal of Educational Psychology*, *102*, 257–280.

Common Core Standards Initiative. (2012). Retrieved from http://www.corestandards.org/assets/CCSSI_ELA%20Standards.pdf

Connor, C. M., Morrison, F. J., Fishman, B., Guiliani, S., Luck, M., Underwood, P. S., . . . Schatschneider, C. (2011). Testing the impact of Child Characteristics × Instruction interactions on third graders' reading comprehension by differentiating literacy instruction. *Reading Research Quarterly*, *46*, 189–221.

Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika*, *16*, 297–334. doi:10.1007/BF02310555

Crooks, T. J. (1988). The impact of classroom evaluation practices on students. *Review of Educational Research*, *58*, 438–481.

Crothers, E. J. (1972). Memory structure and the recall of discourse. In J. B. Carroll & R. O. Freedle (Eds.), *Language comprehension and the acquisition of knowledge* (pp. 247–283). Washington, DC: V. H. Winston & Sons.

Drummond, K., Chinen, K., Duncan, T. G., Miller, H. R., Fryer, L., Zmach, C., & Culp, K. (2011). *Impact of the thinking reader software program on Grade 6 reading vocabulary, comprehension, strategies, and motivation* (NCEE 2010–4035). Washington, DC: National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education. Retrieved from http://ies.ed.gov/ncee/edlabs/regions/northeast/pdf/REL_20104035.pdf

Frederiksen, C. H. (1975). Representing logical and semantic structures of knowledge acquired from discourse. *Cognitive Psychology*, *7*, 371–458.

Gambrell, L. B., Palmer, B. M., Codling, R. M., & Mazzoni, S. A. (1996). Assessing motivation to read. *Reading Teacher*, *49*, 518–533.

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

Graesser, A. C., Lu, S., Jackson, G. T., Mitchell, H. H., Ventura, M., Olney, A., & Louwerse, M. M. (2004). AutoTutor: A tutor with dialogue in natural language. *Behavior Research Methods, Instruments, and Computers*, *36*, 180–192.

Grimes, J. E. (1975). *The thread of discourse*. The Hague, the Netherlands: Mouton.

Guthrie, J. T., & Davis, M. H. (2003). Motivating struggling readers in middle school through and engagement model of classroom practice. *Reading & Writing Quarterly*, *19*, 59–85.

Halliday, M. A. K., & Hasan, R. (1976). *Cohesion in English*. London, UK: 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 edition* (pp. 635–654). Mahwah, NJ: Erlbaum.

James-Burdumy, S., Mansfield, W., Deke, J., Carey, N., Lugo-Gil, J., Hershey, A., . . . Faddis, B. (2009). *Effectiveness of selected supplemental reading comprehension interventions: Impacts on a first cohort of fifth-grade students* (NCEE 2009-4032). Washington, DC: National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education. Retrieved from http://www.mathematica-mpr.com/publications/pdfs/education/selectsupplreading.pdf

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

Krauss, I. K., & Hoyer, W. J. (1984). High technology and the older person: Age, sex and experience as moderators of attitudes towards computers. In P. K. Robinson, J. Livingston, & J. E. Birren (Eds.), *Aging and technology* (pp. 349–350). New York, NY: Plenum.

Lorch, R. F., Jr., & Lorch, E. P. (1995). Effects of organizational signals on text- processing strategies. *Journal of Educational Psychology*, *87*, 537–544.

Mayer, R. E. (2001). *Multimedia learning*. New York, NY: Cambridge University Press.

Meyer, B. J. F. (1971). *Idea units recalled from prose in relation to their position in the logical structure, importance, stability, and order in the passage* (Unpublished master's thesis). Cornell University, Ithaca, New York.

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.), *Understanding expository text: A theoretical and practical handbook for analyzing explanatory text* (pp. 11–64, 269–304). Hillsdale, NJ: Erlbaum.

Meyer, B. J. F. (2003). Text coherence and readability. *Topics in Language Disorders*, *23*, 204–224.

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., & McConkie, G. W. (1973). What is recalled after hearing a passage? *Journal of Educational Psychology*, *65*, 109–117.

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 using 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., & 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, NY: Cambridge University Press.

Meyer, B. J. F., & Wijekumar, K. (2007). A web-based tutoring system for the structure strategy: Theoretical background, design, and findings. In D. S. McNamara (Ed.), *Reading comprehension strategies: Theories, interventions, and technologies* (pp. 347–375). Mahwah, NJ: Lawrence Erlbaum Associates.

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

Meyer, B. J. F., Wijekumar, K., Middlemiss, W., Higley, K., Lei, P., Meier, C., & Spielvogel, J. (2010). Web-based tutoring of the structure strategy with or without elaborated feedback or choice for fifth- and seventh-grade readers. *Reading Research Quarterly*, *41*, 62–92.

National Center for Education Statistics. (2011). *The nation's report card: Findings in brief. Reading and mathematics 2011*. Washington, DC: Author. Retrieved from http://nces.ed.gov/nationsreportcard/pdf/main2011/2012459.pdf

Ponce, H. R., Lopez, M. J., & Mayer, R. E. (2012). Instructional effectiveness of a computer-supported program for teaching reading comprehension strategies. *Computers & Education*, 59, 1170–1183.

Pressley, M. (2002). What should comprehension instruction be the instruction of? In M. L. Kamil, P. B. Mosenthal, P. D. Pearson, & R. Barr (Eds.), *Handbook of reading research* (Vol. 3, pp. 545–561). Mahwah, NJ: Erlbaum.

Pressley, M., Wharton-McDonald, R., Mistretta-Hampston, J., & Echevarria, M., (1998). Literacy instruction in 10 fourth- and fifth-grade classrooms in upstate New York. *Scientific Studies of Reading*, *2*(2), 159–194.

Proctor, C. P., Dalton, B., & Grisham, D. L. (2007). Scaffolding English language learners and struggling readers in a universal literacy environment with embedded strategy instruction and vocabulary support. *Journal of Literacy Research*, *39*, 71–93.

Raudenbush, S. W., & Bryk, A. S. (2002). *Hieararchical linear models: Applications and data analysis methods*. Thousand Oaks, CA: Sage.

Sherer, M., Maddux, J. E., Mercandante, B., Prentice-Dunn, S., Jacobs, B., & Rogers, R. W. (1982). The self-efficacy scale: Construction and validation. *Psychological Reports*, *51*, 663–671.

Slavin, R.E., Cheung, A., Gross, C., & Lake, C. (2008). Effective reading programs for middle and high schools: A best-evidence synthesis. *Reading Research Quarterly*, *43*, 290–322.

Stanovich, K. E. (1986). Matthew effects in reading: Some consequences of individual differences in the acquisition of literacy. *Reading Research Quarterly*, *21*, 360–407.

Theodorou, E. (2006). *Comparing the effects of learning the structure strategy via web-based training or classroom training on the recall of near and far transfer texts* (Unpublished doctoral dissertation). University Park, PA: The Pennsylvania State University.

U.S. Department of Education (2008). *Common Core Dataset*. Retrieved from http://nces.ed.gov/ccd/

van den Broek, P., Rupp, D. N., & Kendeou, P. (2005). Integrating memory-based and constructionists processes in accounts of reading comprehension. *Discourse Processes*, *39*, 299–316.

Wiederholt, J. L., & Blalock, G. (2000). *Gray Silent Reading Tests*. Austin, TX: PRO-ED.

Wijekumar, 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.

Wilson, V. L., & Rupley, W. H. (1997). A structural equation model for reading based on background, phonemic, and strategy knowledge. *Scientific Studies of Reading*, *1*, 45–63.