Education Tech Research Dev (2012) 60:987–1013
DOI 10.1007/s11423-012-9263-4

RESEARCH ARTICLE

Large-scale randomized controlled trial with 4th graders using intelligent tutoring of the structure strategy to improve nonfiction reading comprehension

Kausalai Kay Wijekumar · Bonnie J. F. Meyer · Puiwa Lei

Published online: 26 June 2012

Abstract Reading comprehension is a challenge for K-12 learners and adults. Nonfiction texts, such as expository texts that inform and explain, are particularly challenging and vital for students' understanding because of their frequent use in formal schooling (e.g., textbooks) as well as everyday life (e.g., newspapers, magazines, and medical information). The structure strategy is explicit instruction about how to strategically use knowledge about text structures for encoding and retrieval of information from nonfiction and has consistently shown significant improvements in reading comprehension. We present the delivery of the structure strategy using a web-based intelligent tutoring system (ITSS) that has the potential to offer consistent modeling, practice tasks, assessment, and feedback to the learner. Finally, we report on statistically significant findings from a large scale randomized controlled efficacy trial with rural and suburban 4th-grade students using ITSS.

Keywords Reading comprehension · Intelligent tutoring systems · Web-based learning · Randomized controlled experiment

Introduction

Reading and comprehending texts is an essential part of life today. Whether it is classroom learning, reading the newspaper, or researching a healthcare choice, they all require people to read texts, process the information, integrate the information with prior knowledge, create logical schemas for the text, and recall them when needed in a flexible manner that is conducive to the task at hand. Comprehension is also essential for all content area learning. For example, mathematics word problems require the learner to extract the most important aspects of the problem before they can apply their mathematics skills to solve it.

K. K. Wijekumar (✉)
College of Information Sciences and Technology, The Pennsylvania State University,
212 Ross Administration Building, Penn State Beaver, Monaca, PA 15061, USA
e-mail: kxw190@psu.edu

B. J. F. Meyer · P. Lei
The Pennsylvania State University, University Park, PA, USA

When reading a physics textbook, students must again identify the goal of the reading, identify important concepts and relationships, and be able to apply these concepts when they work on their laboratory activities. Similarly, outside the school environment comprehension of expository text is essential to important everyday activities. For example, when making a healthcare choice, people must read research studies, articles in newspapers, and brochures on the medical condition and treatment options (Meyer and Poon 2001). Again, reading comprehension is a critical component of the health decision-making process.

Unfortunately, U.S. school children have been under prepared for reading and comprehending expository texts as evidenced by their performance on comprehension tests at State and National levels (NAEP 2007). Solutions to the reading comprehension problems in all grades have relied on underlining, summarizing, questioning, building background knowledge, and vocabulary instruction in language arts curricular. Curricula that have been taught for the past 40 years with these strategies have shown a 5-point improvement in reading comprehension measured in grades 4, 8, and 11 by the NAEP (1990, 2007) since 1969. Additionally, James-Burdumy et al. (2009) conducted a year-long large scale randomized controlled trial testing the effectiveness of four popular language arts programs in elementary schools (Project CRISS, ReadAbout by Scholastic, Read for Real by Zaner-Bloser, and Reading for Knowledge by Success for All) and reported no significant differences between all the interventions and control classrooms.

Fourth grade is the youngest grade level tested for reading comprehension in the National Assessment of Educational Progress (NAEP) and shows that 36 % of students score below levels of basic proficiency (NAEP 2007). Fourth grade is also a critical time in transitioning students from story-based reading to more expository text (content area reading) comprehension. Failing to make the transition from narrative to expository text comprehension can result in serious negative effects on subsequent school learning. Chall et al. (1990) report on a fourth-grade slump in reading achievement that may be caused by a lack of skills in reading and comprehending expository texts. Children who do not successfully understand and use comprehension strategies to read and comprehend their textbooks, newspapers, and other content area materials will face increasing challenges in all their academic work as they proceed to higher grades. We selected fourth grade for this research study because of the significance of this year in a child's progress from elementary narrative texts to comprehending expository texts in all their school subjects.

This paper focuses on the structure strategy as a promising approach that has the potential to improve reading comprehension based on findings from many previous studies (e.g., Meyer et al. 1980, 2002, 2010, 2011; Williams et al. 2005, 2009). The structure strategy focuses on common patterns used by writers to organize expository texts and to convey main ideas. These patterns build on one another to convey the logical structure of a text. The structure strategy explicitly teaches students how to follow the logical structure in nonfiction through strategic use of knowledge about text structures. Students learn how to use these structures to increase comprehension and organize their writing about what they remember from reading. For example, students learn about certain vocabulary words, called signaling words ("solution," "in contrast"), that can clue readers into arguments often made in expository text. The structure strategy showed strong effects on understanding everyday texts (e.g., informative financial and medical texts) in a randomized control study in a traditional classroom setting with adult learners (Meyer and Poon 2001). All extant studies, except 5th and 7th grade studies reported by Meyer et al. (2002, 2010, 2011), relied on trained teachers to deliver the structure strategy instruction to learners.

Expanding access to the structure strategy required training large numbers of teachers and providing grade appropriate readings from many different domains to be incorporated into all existing language arts programs. This project focused on the delivery of the structure strategy using a web-based intelligent tutoring system (ITSS) that has the potential to offer consistent modeling, practice tasks, assessment, feedback, and application of the structure strategy to many different domains for the learner. The system was designed to deliver instruction within existing language arts curricular for one class period a week and provide one on one tutoring to students in fourth grade.

The ITSS delivery system was designed based on principles of multi-media learning proposed and investigated by Mayer (2009), particularly the coherence principle and the signaling principle. These principles explain that people learn better without distracting material extraneous to the primary instructional goals and with emphasis of critical words. These principles take into account the limitations of the human information processing system (Baddeley 1986; Sweller 2005). We provided a simple, uncluttered interface and highlighted signaling words, following Mayer's coherence and signaling principles. Extensive details about multi-media research and design decisions for ITSS are provided in Meyer and Wijekumar (2007) also see Meyer et al. (2010). Further description will be presented in the description of the ITSS system and its use with fourth graders.

This research study is also the largest scale randomized controlled trial on both the structure strategy as well as its delivery via a web-based ITSS. It is also important to note that randomized controlled trials in schools settings provide information on the efficacy of such tools in a natural setting with practical implications for schools. Large scale studies such as these are also one approach to draw causal conclusions linking interventions and learning outcomes such as reading comprehension.

Finally, we report on statistically significant findings from a large scale randomized controlled trial with rural and suburban 4th-grade students using ITSS. The structure strategy, design and delivery of ITSS, research design, measures, procedures, and results are described next.

Structure strategy as a tool to solve the 4th-grade comprehension problem

The structure strategy is designed to help readers use signals for text structures in non-fiction to create strategically organized and efficient mental representations and use that knowledge to apply their memory of the text when needed. The three basic tenets of the structure strategy are the signaling words that focus the readers' attention to the text structure organizing the reading passage, then using the text structure to strategically organize the ideas in the passage and create a main idea, and finally creating a well-organized mental representation and recall of the text using the structure. Users of the structure strategy learn how to identify signaling words in a passage. For example, Fig. 1 shows a passage on crocodilians organized using the comparison text structure. Signaling words for the comparison text structure in the crocodilians passage include "different," "differ," and "in contrast." Once the user identifies the signaling words they can classify the passage into one of the five text structures identified by Meyer (1975). These text structures include comparison (sometimes referred to as compare/contrast), problem and solution, cause and effect, sequence, and description.

Each text structure has a unique organization for its main ideas (summary) as well as structuring its recall. In the example provided in Fig. 1, there are two paragraphs each containing a comparison signaling word. A main idea for this passage would be, "alligators and crocodiles were compared on their snouts, teeth, and how long they live." The abstracted pattern for a comparison main idea is ____ and _____ were compared on ____, _____, and _____. With the signaling word and main idea established, the user is able to write two parallel paragraphs using correct signaling words for the comparison structure and all the details necessary to make the passage complete. In this case, the user must write the three pieces of information about both crocodiles and alligators.

In summary, the structure strategy has three basic steps:

1. Identify the signaling word(s) in a passage and classify the text structure. This focuses attention toward the top-level structure and formulating memory representations in a strategic manner.
2. Write a thorough main idea for the passage using the main idea pattern for that text structure. This focuses the learner's attention to the most important ideas in the text that will be at a higher level in the strategic memory representation.
3. Create a recall of the passage using signaling words and the main idea to prompt a complete recollection and construction that is logically organized. This focuses the learner's attention to comprehension monitoring and allows the learner to review their comprehension of the text.

Research about the structure strategy spans 40 years and traces its underpinnings to the organization of prose and how prose structure affects memory representations of text (Meyer and McConkie 1973). Through many years of research with many different groups of users the structure strategy has consistently shown improvements in reading comprehension. For example, Meyer et al. (1989) carried out the earliest randomized control trial showing larger statistically significant gains in reading comprehension for young and older adults taught to use the structure strategy in comparison to those who were randomly assigned to practice reading and recalling the same instructional materials (everyday nonfiction) without structure strategy instruction or a no contact control group. More recently, Meyer et al. (2010) studied the effects of the ITSS structure strategy delivery to 5th and 7th grade students and showed statistically significant improvements in reading comprehension. Similarly, Williams et al. (2005, 2009) have taught at-risk students in second grade the structure strategy and shown statistically significant improvements in their reading comprehension. The current research is the first extension of the structure strategy to 4th-grade students.

The structure strategy can be used as a component of the language arts curriculum and complement all existing language arts practices such as underlining and summarizing. Both underlining and summarizing require the student to be able understand what is important and how the important ideas should be logically organized/related. Using the structure

Crocodilians

Two different kinds of crocodilians exist today; these two types are crocodiles and alligators. These interesting creatures differ dramatically in their snouts, teeth, and how long they live. Crocodiles have long tapered snouts. Their lower teeth are visible sticking out of their closed jaws. Crocodiles live 45 to 50 years.

In contrast to crocodiles, alligators have broad rounded snouts. Their lower teeth are hidden tucked in their closed jaws. Alligators live 30 to 40 years.

Fig. 1 Comparison text structure passage about crocodilians strategy allows students to easily identify important parts of the passage. The structure strategy's main idea patterns give them even more help in identifying and strategically using the organization of the text. Because the structure strategy can work jointly with other comprehension strategies the training for the method can be delivered as a supplement to the language arts curriculum. This research study presented the structure strategy training as a partial substitute for the language arts curriculum for 30–45 min each week over six months.

Another important aspect of the structure strategy is that it applies to any domain. For example, the comparison text structure can be used to organize a passage on science topics, such as crocodilians, and the same text structure can be used to compare three Olympic athletes in a sports article. The problem and solution text structure can be used to present an article on the declining populations of sharks in the oceans and also used to organize an article on problems with economic policies. It is important to note that students learning how to use the structure strategy must be explicitly taught how to transfer their learning from one domain to another as observed by Theodorou (2005). Based on this finding the structure strategy delivered in this study presented passages from science, social studies, sports, and current events. Students also learned how to use the five different text structures within these domains.

The structure strategy further shows learners how to combine text structures to form complex passages. For example, students can be reading a passage that discussed the oil spill in the gulf. This passage can be organized as a problem and solution. In this case, the problem is that there is a massive oil spill in the gulf. There is more than one solution to the problem requiring the author to have a comparison of the possible solutions. So the article will have an overall organization using the problem and solution text structure. The comparison of the possible solutions, i.e., top-kill or bottom-kill the well can be organized using a comparison text structure.

In summary, the structure strategy shows students how to strategically organize their memory of expository texts and works to enhance other comprehension approaches, such as underlining and summarizing, by helping students find the most important information (Meyer et al. 1989). The structure strategy can be applied to any domain and text structures can be nested to form complex passages. Unfortunately, the biggest drawback to delivering the structure strategy to a larger audience was the need for trained teachers and training materials. Additionally, the students needed to learn the strategy and apply the approach in different domains within a limited timeframe. One-to-one tutoring is a solution to addressing this need and a web-based ITSS appeared to be an efficient and effective solution based on preliminary work by the authors in grades 5 and 7 (Meyer et al. 2002, 2010).

Design and delivery of ITSS

The ITSS system was designed using theories of multi-media learning and core instructional design principles focusing on the learning goals and the necessary conditions, activities, and assessments designed to achieve the goal. The learning goal for ITSS was for students to understand and apply the structure strategy in many different domains and with texts of varying difficulty levels. The ultimate goal was improved reading comprehension of expository texts.

The advantage of a web-based delivery method is the consistency of modeling, practice tasks, assessments, feedback (Anderson et al. 1995; Meyer and Wijekumar 2007), and individual self-paced learning for the student. ITSS uses an animated pedagogical agent (APA) tutor, called I.T. (standing for Intelligent Tutor and shown in Fig. 2), as the expert who helps and guides the learner. I.T. models how he would find signaling words in a passage, identify the text structure, construct a main idea for the passage, and write a full recall of the passage. Each sub-goal (e.g., identify the signaling word) is carefully crafted with I.T. introducing the approach, showing the student how to find the signaling word (I.T. speaks as the signaling word blinks on the screen), and then allowing the student to practice finding the signaling word themselves. After I.T. completes the modeling, the student is allowed to click on the signaling words they find in the passage. The student is allowed multiple attempts to answer the question with progressively more help to solve the problem. After the first attempt, the student may be asked to "try again." During the second try, the student may be shown a table of signaling words with I.T. prompting them to read the words and try to find the correct words in the passage. If the student has not mastered the task after the fourth or fifth try, they will be prompted with the correct answer and transitioned into another similar passage. Students who experience continued difficulty in mastering the concept are prompted to request help from their teacher.

Sub-goals within the system include identifying signaling words, classifying the text structure, writing a main idea, writing a full recall, and variations to these goals designed to motivate the students. For example, in some lessons students write a main idea but in other lessons they click and fill in a table for the main idea (see Fig. 3). Another approach used to represent the main ideas is a tree-like presentation showing the ideas being compared (see Fig. 4). This type of variation in presentation allows the students to experience the different representations of the logical organization of the text.

Students make progress through the system by listening to I.T.s modeling and instructions followed by completing practice tasks designed to help the student achieve the

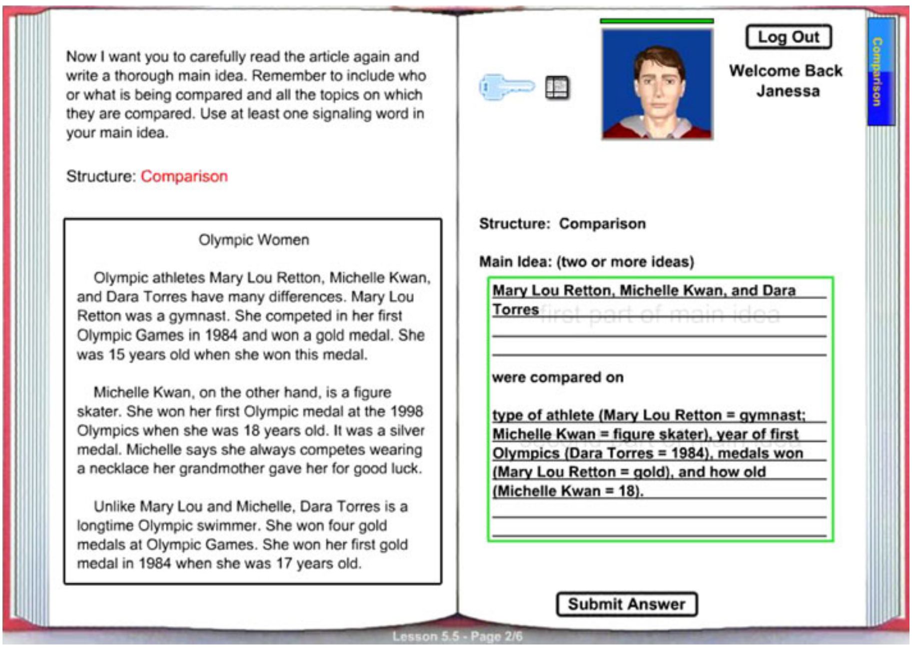

Fig. 2 ITSS book-like interface with APA I.T

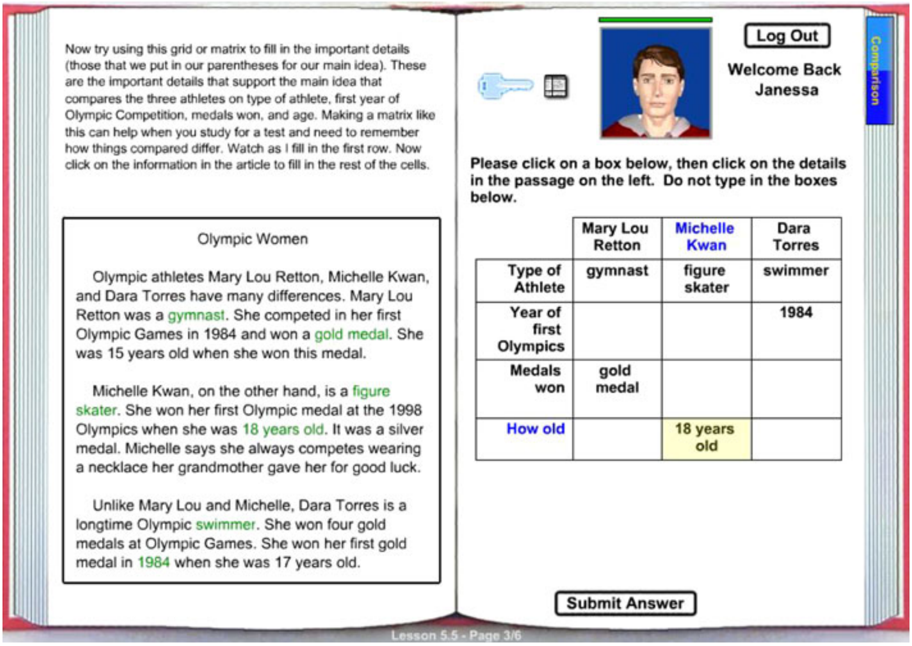

Fig. 3 ITSS—matrix display of main idea

intermediate learning goals. Each task has been carefully crafted with differing number of tries and scaffolding (e.g., providing an initial comparative sentence for the recall task in an early lesson). The numbers of tries are varied to disallow students' gaming of the system. If a student notices that he/she is provided the correct answer after every third try, then he/she may be tempted to just submit nonsense answers initially and wait to get the correct answer in the third try. Instead, when the system varies the numbers of tries students are kept guessing and cannot "game" the system. The scaffolding is initially designed to give students needed support in understanding how to apply the concepts. As students gain confidence in completing the tasks, the scaffolding is faded away.

The design-goal for ITSS included setting the tone for reading (with a book-like interface), presenting a "tutor" to interact with the learner playing the role of a teacher, and building interactions between the tutor and student to help the student learn and apply the structure strategy (by modeling, providing practice tasks, assessments, feedback, and scaffolding). The modeling, practice, assessment, and feedback within ITSS was created using a web-based interface, multi-media tools, database driven rules engine, and a.NET and C# framework to deliver the structure strategy training. The design features have been described extensively in Meyer and Wijekumar (2007) as well as Meyer et al. (2010). Next we describe briefly some of the multi-media design features relevant to this research study—affordances, interactions between agent and student, and gaming.

ITSS is designed using many of the multi-media learning principles researched and compiled by Mayer (2009). ITSS was designed to present a book-like interface (Meyer and Wijekumar 2007). First, the human voice, used by I.T. in all the narrations of the modeling, instructions, and feedback, fosters social agency (Atkinson et al. 2005). Atkinson et al. reported that students relate better to a human voice over computer generated voices and ITSS

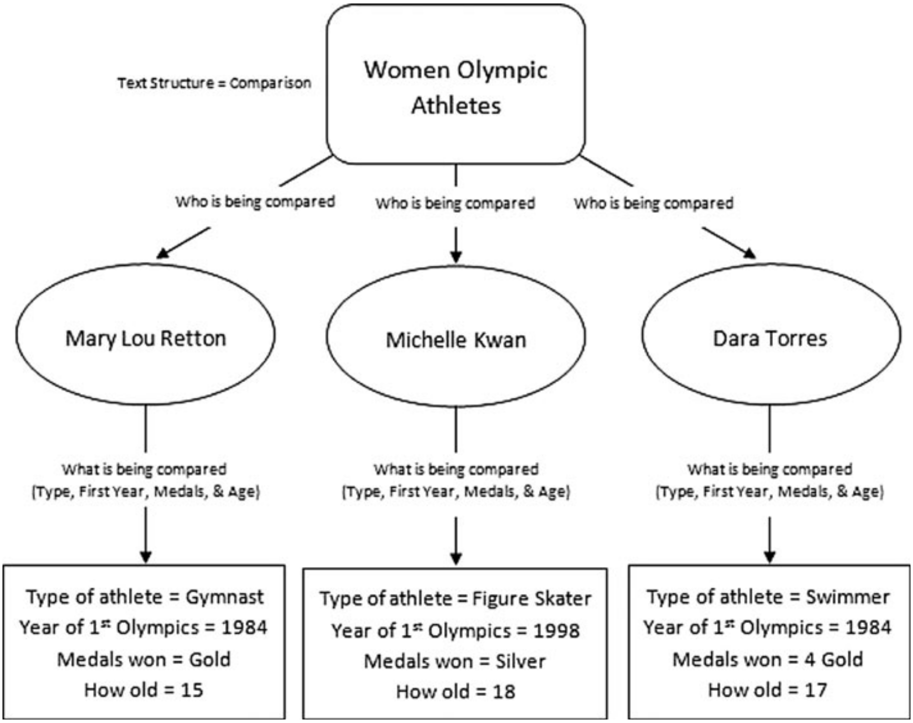

Fig. 4 ITSS—tree style presentation of main idea

uses a human voice rendered to match the APA's facial movements. Our system also minimizes the seductive details (Mayer's Coherence Principle) that can hurt the learning process (Harp and Mayer 1998). Harp and Mayer noted that students with low prior knowledge were easily distracted by images and unrelated graphics that they referred to as "seductive details". Within ITSS, instead of animated cartoonish characters that are frequently used in multimedia learning environments, our system focuses the students' attention to the important aspects of the text and the APA, I.T. acting as a "teacher" (Meyer et al. 2002).

The system also uses the narration with the screen presentation of the text allowing the reader to view the text while listening to the narration. This approach is slightly different than Mayer's (2009) conception of the modality principle. Our approach allows the learner to read the text (using their visual channel) and listen to the narration (using their auditory channel). In the case of younger children in 4th grade this is especially helpful for students who are reading unfamiliar words. Another advantage of these narrations in ITSS is that the student is not able to move to the next page until I.T. has completed speaking. The navigation buttons do not appear on the screen until the animation is completely viewed. This stops students from skipping over pages, a problem found with fifth graders in our initial web-based approach (Meyer et al. 2002) as well as other multi-media instruction (Jacobson et al. 1995). To minimize any frustration that fluent readers may experience because of the narration, teachers have the ability within ITSS to turn off narrations for proficient students. However, this option was not activated for teachers in the current study. This decision was made because of our younger participants, where less reading fluency was expected with fourth graders encountering expository text, and design trade-offs to keep the experimental condition similar among the many school districts.

Students using ITSS receive immediate feedback from the system based on their responses. The student responses are parsed and checked against the scoring tree stored as a linked list in the ITSS database. If there are no matches against the scoring tree additional processing is performed to check spelling and synonyms. ITSS is capable of using latent semantic indexing for comparing student responses with the expert models. The system is also designed to store student responses and create new pathways through the system as new responses are added to the system. The system stores scores for the signaling words, main idea pattern, details, and the text structure used. The ITSS feedback system selects appropriate responses from the database based on the try and score combinations. Additionally, as mentioned previously, the number of possible tries is varied to prevent students from gaming the system.

Figure 5 shows the ITSS interaction cycle beginning with I.T. initiating the tutoring with information for the learner. I.T. also models how to read and look for signaling words and asks the student to practice doing the same. The interactions continue until the student has completed writing a main idea and full recall for the passage. The main idea is typed by the student while the passage is visible on the screen. The full recall, however, is only completed after the student has carefully read the passage and then pressed the "finished reading" button. At that point, the passage is removed from the screen and the student has to write his/her recall based on his/her memory of the text. Students are also advised to use their signaling words and reminded to check for the correct logical organization of the passage before I.T. proceeds to assess their response.

An important modification to the ITSS program was incorporated after the first month of usage during the research data collection. Fourth grade teachers contacted the research team and stated that most 4th grade students were in the "hunt and peck" stage of typing and were having difficulty typing the full recall task. This was very frustrating to the students and teachers. After discussions within the team the ITSS software was modified for 4th-grade classrooms allowing students to complete the clicking on signaling words task and the writing of the main idea tasks. That is, the signaling word identification and writing a main idea tasks were not modified for this study. The full recall task was removed from the lesson sequence for all of the 4th-grade ITSS instruction. In this implementation, students would read the passage, identify the signaling words, write a main idea, receive feedback on both those tasks and then move to the next lesson.

Finally, we refer to ITSS as intelligent because the assessment system uses approaches ranging from keyword matching, comparing responses to expert responses, latent semantic indexing, and continuously updates pathways through the stored responses to provide accurate feedback to the learner. The latent semantic indexing scoring was compared to the former two approaches and found to be lacking accuracy and inefficient because of the short texts used in ITSS. Additionally, students within the ITSS system write main ideas and recalls for pre-determined passages and do not write creative pieces with open-ended questions. Therefore, the ITSS system used in this implementation uses keyword matching and comparing responses to expert texts for the assessment and feedback.

Research questions

The primary research question was: Do 4th grade classrooms using the ITSS system as a partial substitute for the standard language arts curriculum outperform control classrooms on standardized and researcher-designed measures of reading comprehension?

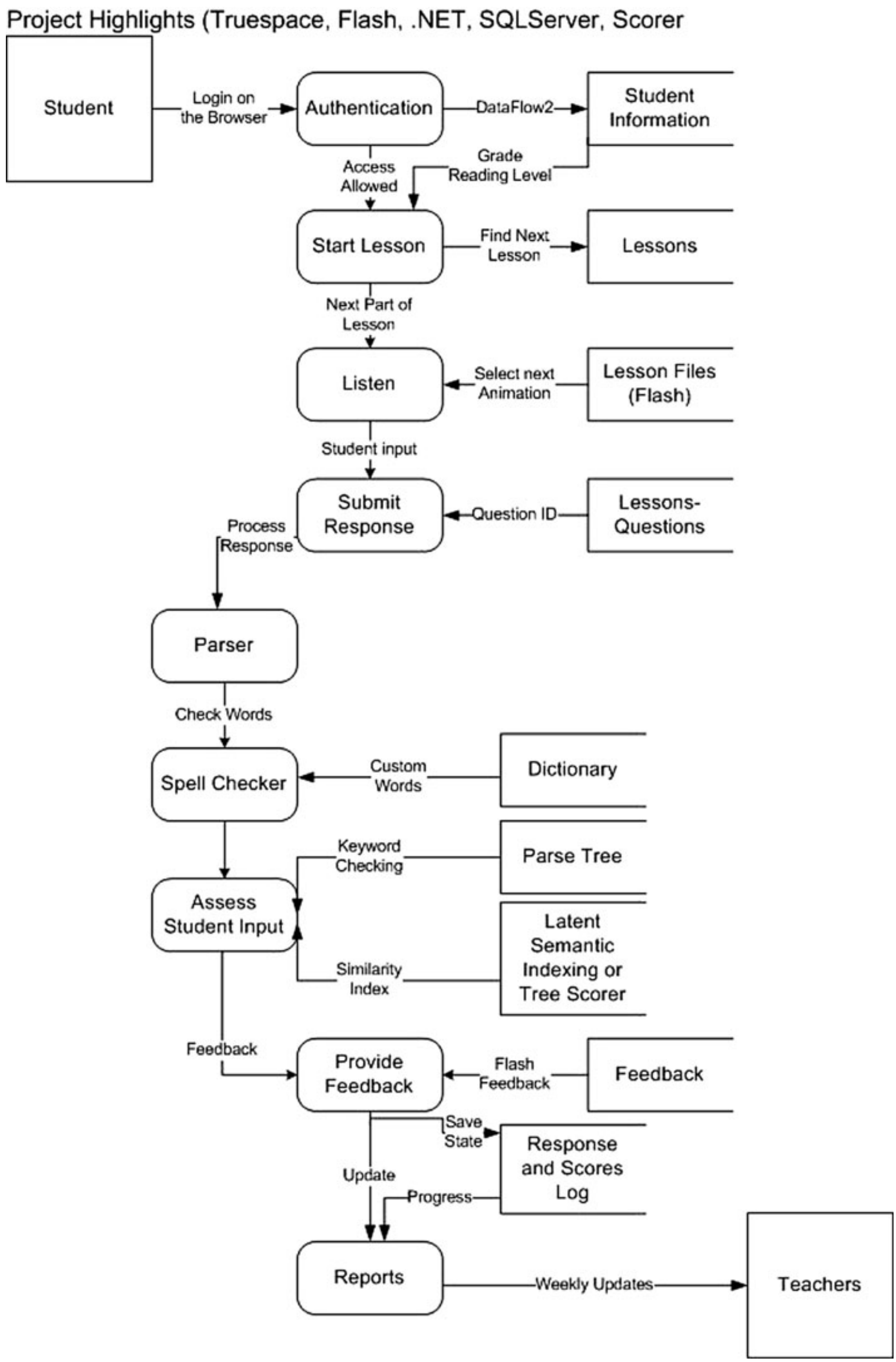

Fig. 5 ITSS interaction cycle

The study also posed six exploratory 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. The six secondary questions were: Does the effect of ITSS on reading comprehension differ between male and female students? This question was posed because research (e.g., Halpern 2006) indicates that boys have more difficulty with verbal tasks than girls, particularly those involving writing. The next two exploratory questions were posed because there is controversy in the literature about initial reading comprehension proficiency and how these differences play out over time or with treatments (e.g., Stanovich 1986; Shaywitz et al. 1995). 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? The following two questions focus on the school contexts and relate to the design factors in the study; achievement risks have been noted for children attending rural schools in PA (House Commission on Rural Education (HCRE) 2004). 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? The final question relates to frequency of use/attendance at ITSS sessions; amount of treatment has been related to treatment effects for other reading interventions (e.g., Caccamise et al. 2010). Do students who used the ITSS system more perform better on the post test than students who used it less?

Based on the instructional goals and the design of the ITSS system we hypothesized that 4th-grade students using the ITSS system for one class period a week for the academic year will outperform their control counterparts on standardized and researcher designed measures of reading comprehension.

Method

Research design

We used a *Multi-Site Cluster Randomized Trial* (*CRT*) design to test the efficacy of ITSS with 4th-grade rural and suburban students as shown in Fig. 6. One hundred and thirty-one teachers' classrooms were randomly assigned to experimental conditions (ITSS and control) within schools. If a school did not have enough classrooms, we grouped schools with similar characteristics to form a site before random assignment of classrooms to treatment conditions. A statistical power analysis, which assumed a minimum detectable effect size (ES) of .20, showed that a minimum of 56 rural and 56 suburban classrooms would be needed for the study.

The final pool of participants included 60 rural and 71 suburban classrooms. All schools volunteered to participate in the study and were not randomly sampled from the universe of eligible schools in the region. Schools signed a memorandum of understanding agreeing to participate in the study and also to allow classroom visits from the research team and provide the necessary computer laboratory time to complete the research study.

The multisite CRT design that uses teacher random assignment, within each school, was selected over other designs that use school- or student-level random assignment. A design based on student-level random assignment was considered but rejected because of the expectation that school officials, teachers, and parents would object to leaving student placement in classrooms to chance, creating challenges to school recruitment. Additionally, a student level random assignment would have resulted in both intervention and control

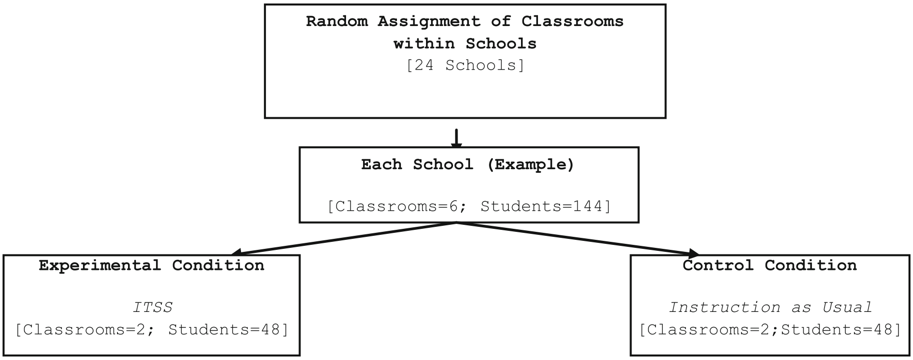

Fig. 6 Random assignment example

students being members of one teacher's classroom resulting in possible contamination of the control group. The contamination could be a result of the teacher professional development (that they would have to withhold information from half their students in the control group) and/or during the computer time when ITSS students were using the software and the rest of the students had to complete their regular language arts curriculum. Furthermore, random assignment of teachers rather than students reflects the software's typical implementation, as well as offering the other advantages. A brief description of additional justifications for choosing the multisite CRT design is presented below.

Statistical power

The statistical power analyses showed the within-school random assignment design to be more efficient than the school random assignment design. Holding constant other assumptions used in a statistical power analysis, the within-school design required approximately half as many schools as the school-level design to detect the same effect.

Curricular consistency between intervention and control

A within-school random assignment, which randomly assigned classrooms within schools to either the intervention or the control group, ensured that the same curriculum was used in both study conditions in each school.

Access to ITSS as a study recruitment tool

This design offered all teachers professional development and the opportunity to eventually use the ITSS software. The intervention teachers received professional development to deliver the instruction in 2009/2010, while the control teachers were offered the same professional development for the following year once the study was completed, along with the option to use ITSS.

Delivery of ITSS and intervention diffusion

Intervention teachers used the ITSS software in their classrooms or in a computer lab in the school. To limit the risk of intervention diffusion (the use of ITSS in control classrooms), the intervention teachers were instructed not to share their software access passwords or professional development materials with other teachers in the school. The expectation of no diffusion of the ITSS intervention to control teachers and their classrooms was reasonable, because control teachers did not receive professional development, and could not view the lesson contents or use ITSS in their classrooms without a password. The risks and consequences of such contamination were explained to teachers and administrators during recruitment and training; classroom observers, who documented the instructional activities in intervention and control schools, were asked to note any apparent use of ITSS in control classrooms.

Finally, students in the control classrooms were given the opportunity to use ITSS for the next academic year. Teachers in the control classrooms were given the full professional development and access to the software as soon as post tests were completed.

Procedure

Within each participating school, all grade 4 teachers' classrooms were randomly assigned to the ITSS intervention or control groups. The control group in each school used the same language arts curriculum as the intervention group in that school. The random assignment produced two groups of classrooms that did not differ significantly on a pre-intervention measure of reading comprehension achievement or other characteristics, including socioeconomic status, percentage of English language learner students, racial/ethnic minority students, and gender.

Teachers in the intervention condition were advised and regularly reminded to use ITSS for 30–45 min each week as a partial substitute for the regular language arts curriculum by the study team. Total time for daily and weekly language arts instruction was to be identical for both the intervention and control classrooms. The time spent on ITSS was expected to be integrated into the overall language arts instructional time to avoid confounding the amount of instructional time with the use of ITSS.

After random assignment the research team visited each school and conducted professional development sessions for teachers assigned to the ITSS group. During those visits all teachers completed the teacher consent forms. Pretests were scheduled at the beginning of the 2009–2010 academic year. Students completed both the *Gray Silent Reading Test* (GSRT) and the researcher designed measures during a large group testing session administered by trained research team personnel with the support of the teachers and school personnel.

After completing the pretests, the research team scheduled the ITSS start-up sessions. Each ITSS teacher was given the complete package of student usernames and teacher manual. Some classrooms were also provided with earphones when necessary. Teachers were instructed to bring their classrooms to the computer lab and hand-out the username sheet to each student. Each student sat at a computer, placed the earphones on his/her ears and opened the web browser to the ITSS site. The student then entered his/her username and password and the ITSS system through I.T. took charge of the learning interactions. The students' interacted with ITSS by listening to I.T., clicking on signaling words, writing a main idea, completing a main idea matrix, and other similar tasks. As noted earlier, this 4th-grade implementation of ITSS did not include the full recall task that is used with other grade levels, and thus, missed a third of the three components for the complete structure strategy training.

After each 30–45 min session (once a week) with the ITSS software, students logged off the computer and returned their username sheets to the teacher. ITSS saved the spot where the learner logged off and placed them at the same spot in the lesson and page when the student returned for the next session.

Fidelity of implementation was monitored by the research team by reviewing the weekly computer usage logs on ITSS. Additionally, the team conducted two fidelity observations during the school year and observed both the intervention and control classrooms. The research team mailed bi-weekly reports to the ITSS teachers on student progress and gaming.

Students used the ITSS software for 6–7 months and then completed the pos*t* tests. The pos*t* tests were completed in similar large group testing sessions administered by the research team and assisted by the teachers and school personnel. Both the GSRT and the researcher designed measures were administered at post-test.

Materials

Reading comprehension was measured using a standardized reading comprehension test and an experimenter-designed recall and main idea tests about expository texts.

Standardized test of reading comprehension

The GSRT (Wiederholt and Blalock 2000) form B was administered at pretest and Form A was administered at post-test. The GSRT is administered in large group settings and contains 15 stories with five questions for each story. The 15 stories get progressively more difficult. 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 responses were entered into a spreadsheet and scored using a computer program with raw scores, top-down and bottom-up scoring methods.

Experimenter-designed measures of reading comprehension

The experimenter measures used in the data analyses included the Signaling test (generating signaling where blanks appeared in comparative text), main idea quality (competent use of the comparison structure writing a 2-sentence main idea), and comparison competency (competent use of the comparison structure from the recall task), and total recall scores from the comparison recall task. The problem and solution measures were problem and solution competency and total recall scores.

These variables were extracted from two equivalent test forms created by Meyer et al. (2010) and one was administered as pretest before the children started ITSS. The second form was administered as post-test immediately after the ITSS use was completed. These measures tested 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 dogs (pretest) or rats (post-test). 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. These measures were transcribed into a spreadsheet and then scored by two trained raters.

Percentage of agreement between two independent scorers for the number of ideas remembered (total recall) on the problem and solution set of texts was 98 %.

A set of two passages also were prepared for the comparison structure: (a) pygmy versus emperor monkeys for the pretest and (b) Adelie versus Emperor Penguins on the post-test. Each comparison text has 128 words, 15 sentences, and 96 idea units. There were three tasks for the comparison structure involving this set of two passages: (a) fill in the blanks where signaling words were removed to gauge signaling word use, called the Signaling test (scores from 0 to 28) (b) a recall task like the recall task used for the problem and solution set of articles, and (c) a comparison main idea task where the student is asked to write a two-sentence main idea with the text available for consultation. Again, these measures were transcribed into a spreadsheet and scored by 2 trained raters. Percentage agreement among scorers for the Signaling test and total recall were 98 and 99 %, respectively.

The scorers for all measures were blind to the experimental condition of the participants. Meyer's (1975, 1985) prose-analysis system was used to score the experimenter-designed measures. At least 10 % of the data from each of the measures were randomly selected from the two conditions (ITSS intervention and control) to calculate inter-rater agreement.

The main idea task was scored for competent use of the comparison structure in writing a main idea (called main idea quality). These quality ratings varied from 1 to 6 depending on how well the reader employed the affordances of comparison structure to organize information from the text about two animals compared on several issues. Main ideas were scored by four graduate students trained extensively using scoring manuals based on Meyer's approach to discourse analysis; percentage of agreement among the four scorers was 93 %. A main idea quality score of 1 indicated no comparison of the two ideas contrasted in the text, while a score of 6 on the scale indicated that the main idea compared two correct creatures on at least two issues [including one super-ordinate issue generated by the student (i.e., "diet"); see Meyer et al. 2010 for the first use of this competency level/main idea quality measure and further description].

Additionally, pairs of trained graduate students scored recalls for competency using the comparison and problem and solution structures on the recall tasks (called comparison competency and problem and solution competency). Scores ranged from 1 to 8 on these scales and percentage agreement between scorers was 89 %. Scores of 1 indicated that the student's recall contained nothing about the text structure (i.e., no correct problem, no correct cause, and no correct solution; no creatures compared). A score of three indicated partial use of the structure (e.g., a correct problem, but no solution), while a score of five indicated good use of the structure and eight indicated mastery (e.g., correct problem and its cause along with the correct solution and blocking of the cause). Examples of student recalls and problem and solution competency scores can be found in Meyer et al. (2010).

These researcher-designed measures have been used in many of the previous studies by Meyer and colleagues (e.g., Meyer et al. 2010, 2011). Reliability of scoring these measures is high in terms of inter-rater agreement, as discussed above, and reliability coefficients, which range from $r = .93$ to .99 (see Meyer et al. 2002). When using these measures students respond to the signaling word task and main idea task while the passage is in view and then write a recall without the passage (they tear the page off and place it in an envelope prior to writing their recall). The measures have good face validity as many of the tasks parallel those performed in classroom activities (e.g., summarizing a reading) and construct validity as they correlate with standardized reading assessments used in many states (Meyer et al. 1980).

Analysis

At posttest the sample included 130 teachers, and 2,643 students, approximately balanced across intervention and control conditions. The analyses tested the mean difference of student achievement between the intervention and control conditions at the classroom level while accounting for students clustered by classrooms, which were clustered by schools.

To determine if there are differences among intervention levels with respect to reading performance outcomes, a series of hierarchical linear modeling (HLM: Raudenbush and Bryk 2002) equations were specified. Analyses were run for each of the primary dependent variables.

HLM model specifications: addressing research questions 1–4

For the HLM models, students were nested in classrooms and classrooms were nested within school districts. In this three-level structure, there are predictor variables at each level. At the student level, predictor variables included reading comprehension covariates (e.g., pretest scores on GSRT and experimenter-designed measures).

The model was specified using Raudenbush and Bryk (2002) nomenclature.

Level 1 (student level)

$$\mathrm{Y}_{ijk} = \pi_{0jk} + \pi_{1jk}\,(female)_{ijk} + \pi_{2jk}\,(GSRT\_pretest)_{ijk} + \pi_{3jk}\,(recall\_pretest)_{ijk} + e_{ijk},$$

where $\mathrm{Y}_{\mathrm{ijk}}$ is the outcome for student i in teacher j's class in school k, π_{0jk} is the average adjusted outcome of students in teacher j's class in school k, π_{1jk} is the difference in outcome scores between male and female students in teacher j's class in school k, *female* is a grand-mean centered indicator variable (1 = female, 0 = male), π_{2jk} is the effect of student level GSRT pretest scores in teacher j's class in school k, *GSRT_pretest* is student's pretest score on GSRT (group-mean centered), π_{3jk} is the effect of student level comparison competency (for recall) pretest scores in teacher j's class in school k, *recall_pretest* is student's comparison competency y pretest score (group-mean centered), e_{ijk} is a random error associated with student i in teacher j's class in school k, and $\mathrm{e}_{ijk} \sim \mathrm{N}\,(0, \sigma^2)$.
The classroom average outcome in a school estimated by the level 1 intercept π_{0jk} was modeled as varying randomly across teachers and as a function of the intervention (partial substitution of ITSS software for regular instruction) at level 2, the teacher level, controlling for the classroom average pretest scores on the GSRT standardized test and comparison competency.

Level 2 (teacher level)

$$\Pi_{0jk} = \beta_{00k} + \beta_{01k}\,(ITSS)_{jk} + \beta_{02k}\,(mGSRT\_pretest)_{jk} + \beta_{03k}\,(mrecall\_pretest)_{jk} + r_{0jk}$$
$$\Pi_{1jk} = \beta_{10k}$$
$$\Pi_{2jk} = \beta_{20k}$$
$$\Pi_{3jk} = \beta_{30k}$$

where β_{00k} is the adjusted average student outcome across all teachers' classrooms in school k, β_{01k} is the adjusted difference in student outcome between the intervention teachers' classrooms and the control teachers' classrooms (intervention effect) in school k, *ITSS* is an effect indicator variable for the intervention that takes a value of 1/2 for an intervention teacher's classroom and −1/2 for a control teacher's classroom, β_{02k} is the effect of the mean classroom GSRT pretest score on classroom average student outcome in school k, *mGSRT_pretest* is class average of GSRT pretest score (grand-mean centered), β_{03k} is the effect of the mean classroom comparison competency pretest score on classroom average student outcome in school k, *mrecall_pretest* is class average of comparison competency pretest score (grand-mean centered), and r_{0jk} is a random error associated with teacher j's classroom in school k on classroom average student outcome $r_{0jk} \sim N(0, \tau_{\pi 00})$.

Level 3 (school level) Assuming that the coefficients for classroom average pretest scores and ITSS effects were homogeneous across schools, their effects were fixed at the school level, as shown in the following specification:

$$\begin{aligned}
\beta_{00k} &= \gamma_{000} + \gamma_{001}(rural)_k + u_{00k} \\
\beta_{01k} &= \gamma_{010} \\
\beta_{02k} &= \gamma_{020} \\
\beta_{03k} &= \gamma_{030} \\
\beta_{10k} &= \gamma_{100} \\
\beta_{20k} &= \gamma_{200} \\
\beta_{30k} &= \gamma_{300}
\end{aligned}$$

where γ_{000} is the adjusted average student outcome across all schools, γ_{001} is the difference between rural and suburban schools in adjusted average student outcome, *rural* is a grand-mean centered indicator variable (1 = rural, 0 = suburban), u_{00k} is a random error associated with school k on adjusted school average student outcome $u_{00k} \sim N(0, \tau_{\beta 00})$, γ_{010} is the average intervention effect across all schools after controlling for differences in pretest scores and holding gender and school locale constant, γ_{020} is the average effect of class-level *GSRT pretest* on student outcome across all schools, γ_{030} is the average effect of class-level *comparison competency pretest* on student outcome across all schools, γ_{100} is the average difference in student outcome between male and female students across all schools, γ_{200} is the average effect of student-level *GSRT pretest* on student outcome across all schools, and γ_{300} is the average effect of student-level *comparison competency pretest* on student outcome across all schools.

Of primary interest among the level 3 coefficients was γ_{010}, which represents the intervention's main effect on the outcome across all schools. A statistically significant positive value of γ_{010} would be reason to reject the null hypothesis of no difference between intervention and control groups in favor of the alternative hypothesis that students in the intervention teachers' classrooms demonstrate higher levels of reading comprehension than do their counterparts in the control teachers' classrooms. HLM6 (Raudenbush et al. 2008) was used to analyze all the multilevel models with the default maximum likelihood estimator for three-level models.

In addition to the statistical significance of the effect of the ITSS intervention, the magnitude of the effect was also expressed in standard deviation units. Specifically, the ES was computed as a standardized mean difference by dividing the adjusted group mean difference (γ_{010}) by the pooled within-group standard deviation of the student-level outcome score (pooled within-group student-level standard deviation on the pretest was used when available).

Results

At baseline the ITSS and control groups did not have any statistically significant differences at the random assignment classroom level ($t128 = |.14|$, $p = .89$ for the GSRT;

t129 = |.09|, $p = .92$ for the Signaling test; t117 = |1.17|, $p = .24$ for Main Idea Quality; t127 = |.33|, $p = .74$ for Comparison Competency). Table 1 presents the GSRT class-level means and standard deviations for the ITSS and control groups at pretest. This indicated that the ITSS and control classrooms were comparable in their reading level before the implementation of the experiment.

Class-level descriptive statistics, statistical test results for ITSS effect from HLM Analyses, and ES on posttests are presented in Tables 1, 2 and 3, respectively, for GSRT, the Comparison Structure, and Problem and Solution Structure. Selected HLM results for different models used to address different research questions are shown in Tables 4 and 5 (each column presents results from a model). Results are discussed by research questions below.

Primary research question

The primary research question was: Do 4th grade classrooms using the ITSS system as a partial substitute for the standard language arts curriculum outperform control classrooms on standardized and researcher designed measures of reading comprehension?

To address this question we used results from HLM model M1 (third column of Tables 4, 5). Students in ITSS classrooms on average scored 1.07 points (or .1 standard deviations, $p > .05$) higher on GSRT adjusted post-test scores (see Table 1) and .78 points higher (or .49 standard deviations, $p < .05$) on main idea quality (see Table 2) than students in control classrooms holding reading pretest scores, gender, and school locale constant. Adjusted post-test scores were also statistically significantly higher for students in ITSS classrooms than their control counterparts on all other researcher measures: Signaling test (see Table 2; adjusted difference = 1.91, ES = .28); comparison total recall (adjusted difference = 1.64, ES = .11); comparison competency (adjusted difference = .41, ES = .18); problem and solution competency (see Table 3; adjusted difference = .28, ES = .13); and problem and solution total recall (see Table 3; adjusted difference = 1.88, ES = .18).

The ES of .10 on the standardized GSRT test was considered small, while the ES of .49 on the main idea quality was considered medium. It should be noted that the 4th-grade version of ITSS used in this study was modified to eliminate the practice tasks for full recall because of teachers' concerns about typing. Instead, students concentrated on clicking on the signaling words and completing the main idea tasks only. Thus, the recall tasks can be considered a transfer task because specific instruction and repeated practice opportunities with elaborated feedback were not provided to the 4th graders; therefore, the ES on the full recall tasks measured by total recall were expected to be small. The results

Table 1 Grade 4 class level means, (SD)s, statistical results, and ES of scaled scores of the GSRT (Wiederholt and Blalock, 2000)

	ITSS			Control			Adjusted difference	*t*-test[a]	ES[b]
	n	*M*	*SD*	*n*	*M*	*SD*			
Pretest	65	22.49	4.17	65	22.39	3.84	–	$t(128) = -.14, p = .89$	–
Posttest	64	29.19	4.37	66	27.86	3.89	1.07	$t(82) = 1.77, p = .08$	.10

[a] Simple *t*-test for ITSS–control difference on Pretest scores and *t*-test for ITSS–control difference on posttest scores adjusted for covariates from HLM model

[b] Difference between ITSS and control groups divided by the student-level pooled *SD* on the pretest

Table 2 Grade 4 class level means, (SD)s, statistical results, and ES of HLM analyses on the comparison structure

Measure	ITSS			Control			Adjusted difference	*t*-test	*ES*[a]
	n	*M*	*SD*	*n*	*M*	*SD*			
Signaling test	65	13.86	5.11	66	11.59	4.55	1.91	$t(82) = 4.48$, $p = .000$	.28
Main idea quality	55	3.20	.57	54	2.44	.49	.78	$t(82) = 9.13$, $p = .000$	.49
Total recall	55	21.17	5.52	54	19.57	4.69	1.64	$t(82) = 2.45$, $p = .02$	.11
Comparison competency	55	3.70	.80	54	3.38	.70	.41	$t(82) = 2.89$, $p = .01$	.18

[a] Difference between ITSS and control groups divided by the student-level pooled standard deviation

Table 3 Grade 4 class level means, (SD)s, statistical results, and ES of HLM analyses on the problem and solution structure

Measure	ITSS ($n = 55$)		Control ($n = 54$)		Adjusted difference	*t*-test	*ES*[a]
	M	*SD*	*M*	*SD*			
Total recall	15.18	2.85	13.48	3.18	1.88	$t(82) = 3.37$, $p = .001$	.18
Problem solution competency	3.08	.62	2.79	.62	.28	$t(82) = 2.58$, $p = .01$	.13

[a] Difference between ITSS and control groups divided by the student-level pooled standard deviation

showed that ES were .11 for comparison total recall and .18 for problem and solution total recall.

The study also posed six exploratory 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. The six secondary questions were:

Six secondary research questions

Does the effect of ITSS on reading comprehension differ between 4th-grade male and female students?

Model M2 (see the fourth column of Tables 4 and 5) addressed the research question on whether the effect of ITSS on reading comprehension differed between male and female students. The effect of ITSS varied as a function of gender on the main idea quality outcome but not on the other measures. Specifically ITSS appeared to make a larger difference on male students than their female counterparts as shown in Fig. 7.

Does the effect of ITSS on reading comprehension differ between low- and medium/high-scoring students on a reading comprehension pretest?

There were no statistically significant interactions between ITSS and students who were categorized as low and medium/high-scoring at the pretest level on any of the outcome measures (e.g., column M3 of Tables 4 and 5). That is, the positive ITSS effects seemed to be the same for low- and medium/high-scoring students on reading comprehension pretests.

Table 4 HLM results on GSRT adjusted post-test scores for grade 4

	M0	M1	M2	M3	M4	M5	M6
	Fixed effects						
Intercept	28.48*** (.47)	28.76*** (.39)	28.76*** (.39)	28.66*** (.38)	28.76*** (.39)	28.80*** (.40)	28.76*** (.39)
Female		−.33 (.50)	−.33 (.50)	−.20 (.54)	−.32 (.50)	−.31 (.50)	−.32 (.49)
Gray pretest[a]		.54*** (.02)	.54*** (.02)	– 7.83*** (.43)	.54*** (.02)	.54*** (.02)	.54*** (.02)
Comparison competency pretest		1.09*** (.13)	1.09*** (.13)	1.55*** (.13)	1.10*** (.13)	1.09*** (.13)	1.09*** (.13)
Class average Gray pretest		.77*** (.09)	.77*** (.09)	.46*** (.10)	.77*** (.09)	.76*** (.08)	.77*** (.09)
Class average comparison competency pretest		.43 (.44)	.43 (.44)	.36 (.44)	.43 (.44)	.54 (.45)	.35 (.44)
Rural		−1.44 (.81)	−1.44 (.81)	−1.13 (.77)	−1.44 (.81)	−1.41 (.82)	−1.39 (.80)
ITSS		1.07 (.60)	1.07 (.60)	.76 (.62)	1.07 (.60)	1.04 (.62)	.92 (.61)
ITSS × female		–	−.10 (.85)	–	–	–	–
ITSS × Gray pretest [a]		–	–	1.30 (.80)	−.01 (.05)	–	–
ITSS × comparison competency pretest		–	–	–	−.15 (.15)	–	–
ITSS × rural		–	–	–	–	−1.83 (1.28)	–
	Random effects (variances of)						
Sites	2.42*	2.73***	2.73***	2.05**	2.73***	2.94***	3.08***
Classrooms	8.64***	2.23**	2.23**	2.37*	2.23**	1.94*	.90
Students	136.99	93.51	93.51	106.69	93.48	93.49	93.51
ITSS	–	–	–	–	–	–	4.18**
	Model fit statistics						
Deviance	18503.42	15712.22	15712.21	15983.25	15711.58	15709.15	15709.35
Number of parameters	4	11	12	12	13	12	13

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

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

Does the effect of ITSS on reading comprehension depend on students' initial reading level?

There is a statistically significant interaction between ITSS and students' reading level (measured by their comparison competency pretest score) on their comparison competency post-test score. Figure 8 presents the pattern of this interaction. ITSS made a larger

Table 5 HLM results on comparison main ideas quality posttest scores for grade 4

	M0	M1	M2	M3	M4	M5	M6
	Fixed effects						
Intercept	2.84*** (.04)	2.85*** (.04)	2.85*** (.04)	2.83*** (.04)	2.85*** (.04)	2.85*** (.04)	2.85*** (.05)
Female	–	.05 (.06)	.06 (.06)	.05 (.06)	.05 (.06)	.05 (.06)	.05 (.06)
Gray pretest[a]	–	.04*** (.00)	.04*** (.00)	– .63*** (.07)	.04*** (.00)	.04*** (.00)	.04*** (.00)
Comparison main idea quality pretest	–	.16*** (.03)	.16*** (.03)	.20*** (.03)	.16*** (.03)	.16*** (.03)	.16*** (.03)
Class average Gray pretest	–	.05*** (.01)	.05*** (.01)	.03** (.01)	.05*** (.01)	.05*** (.01)	.05*** (.01)
Class average comparison main idea quality pretest	–	.13 (.13)	.13 (.13)	.14 (.14)	.13 (.13)	.11 (.13)	.14 (.13)
Rural	–	.01 (.09)	.01 (.09)	.04 (.09)	.01 (.09)	.00 (.09)	.00 (.09)
ITSS	–	.78*** (.09)	.79*** (.09)	.76*** (.09)	.78*** (.09)	.78*** (.08)	.78*** (.09)
ITSS × female	–	–	−.37* (.16)	–		–	–
ITSS × Gray pretest[a]	–	–	–	−.32 (.17)	.01* (.01)	–	–
ITSS × comparison main idea quality pretest	–	–	–	–	.02 (.04)	–	–
ITSS × rural	–	–	–	–	–	.15 (.17)	–
	Random effects (variances of)						
Sites	.000	.000	.000	.000	.000	.000	.004
Classrooms	.28***	.09***	.09***	.10***	.09***	.09***	.09***
Students	2.44	2.14	2.13	2.20	2.14	2.14	2.14
ITSS	–	–	–	–	–	–	.01
	Model fit statistics						
Deviance	7103.43	6257.02	6250.28	6304.80	6253.15	6256.40	6256.96
Number of parameters	4	11	12	12	13	12	13

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

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

difference in comparison competency for students who had a lower initial reading level as measured by comparison competency pretest than those who had a higher initial reading level on this measure.

There is also a statistically significant interaction between ITSS and students' reading level (as measured by the GSRT pretest) on their main idea quality post-test (see column M4 of Table 5). Figure 9 shows that ITSS made a slightly larger difference in main idea quality for students who had higher initial reading levels as measured by the GSRT pretest than students who had lower initial reading levels on the GSRT.

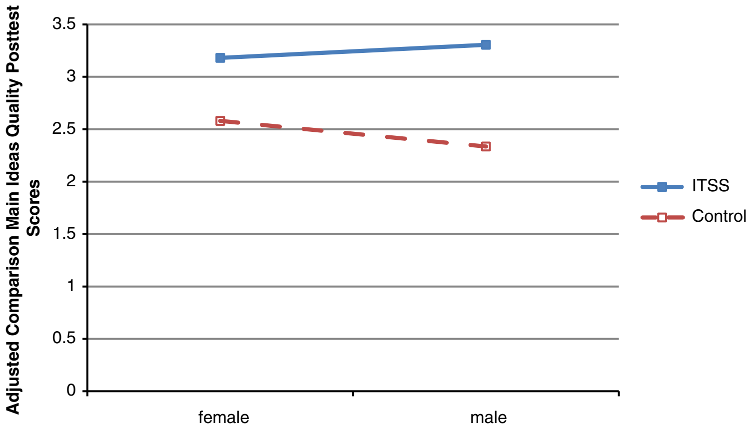

Fig. 7 ITSS × gender interaction on main idea quality

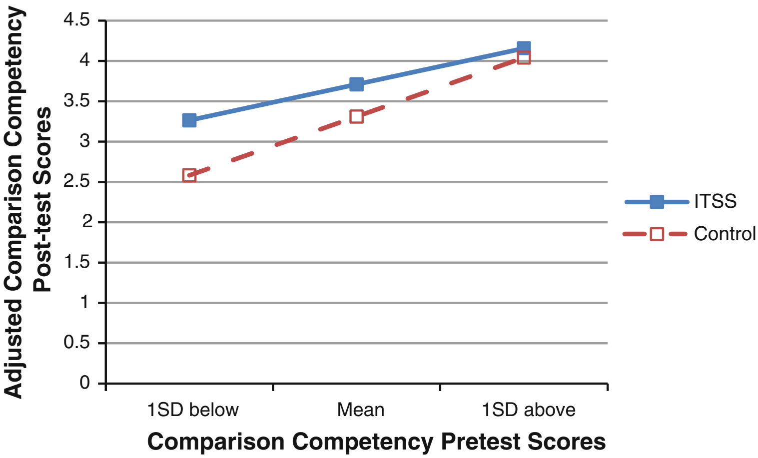

Fig. 8 ITSS × comparison competency pretest interaction on comparison competency post-test

Does the effect of ITSS on reading comprehension vary across rural versus suburban areas?

Model M5 addressed the research question on whether the effect of ITSS on reading comprehension differed between rural and suburban students. There were no statistically significant interactions between ITSS and school locale on any of the outcome measures. The positive ITSS effects were similar for both rural and suburban schools.

Does the effect of ITSS on reading comprehension vary across schools?

Model M6 addressed this research question. The estimated variance of adjusted ITSS effects on GSRT post-test scores was statistically significantly different from zero ($\chi^2 = 50.56$, df $= 29$, $p < .05$). The 95 % plausible value range for adjusted ITSS effects on GSRT post-test scores was (−3.09, 4.93). However, difference in Deviance between the

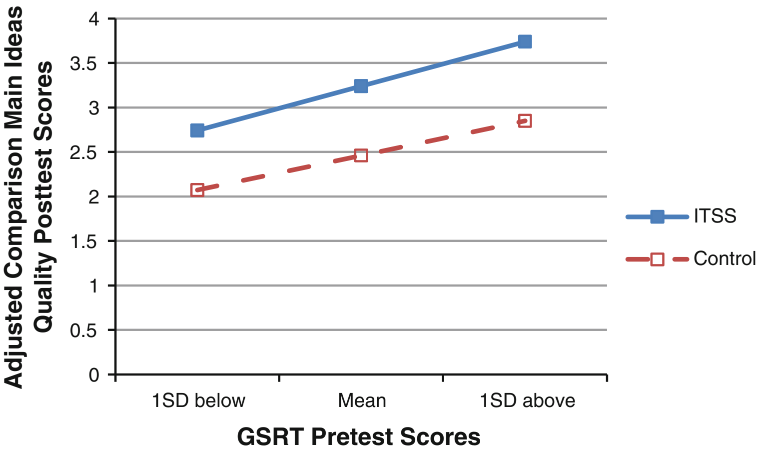

Fig. 9 ITSS × GSRT interaction on main idea quality post-test scores

random ITSS effect model (M6) and the fixed ITSS effect model (M1) on GSRT was not statistically significant (Δ deviance $= 2.87$, $df = 2$, $p > .05$). This suggested that explanatory power was not significantly improved by modeling ITSS effect as random and that the fixed-effect model appeared sufficient (Raudenbush and Bryk 2002).

The effect of ITSS appeared to vary across schools on two measures, the Signaling test and comparison competency. The estimated variances of adjusted ITSS effects on these measures were statistically significantly different from zero ($\chi^2 = 54.98$, $df = 30$, $p < .05$ and Δ deviance $= 8.52$, $df = 2$, $p < .05$ for Signaling test; $\chi^2 = 46.78$, $df = 24$, $p < .05$ and Δ deviance $= 6.52$, $df = 2$, $p < .05$ for comparison competency). The 95 % plausible value range for adjusted ITSS effects on signaling post-test scores was (−1.13, 4.63). The 95 % plausible value range for adjusted ITSS effects on comparison competency post-test scores was (−.63, 1.41). The effect of ITSS did not vary significantly across schools on other outcome measures.

Do students who used the ITSS system more perform better on the post-test than students who used it less?

Pearson correlations of GSRT reading post-test scores with fidelity measures were calculated to address the question of whether students used the ITSS system more performed better on the post-test. Both average minutes used per week and total number of questions answered by students correlated positively with GSRT post-test scores. The strength of the correlation was weak ($r = .09$, $p < .05$) for minutes used per week, but total number of questions answered had a higher and statistically significant correlation with GSRT ($r = .20$, $p < .05$). These correlational analyses suggested that the number of questions answered was a good indicator of fidelity and the total time spent on ITSS also may be used even though the correlation was weaker than the number of questions answered.

Discussion

This study contributes three major findings to the continuing challenge of improving expository reading comprehension that is an essential component to most academic and non-academic activities. First, this research addresses the need to expand the reach of ITSS to a wider audience through the use of web-based ITSSs. As part of designing effective multi-media based learning environments our research focuses on delivering the structure strategy using multi-media learning principles and minimizing seductive details (e.g., Mayer's 2009 principle of coherence). Second, this research study is the first time that Meyer's structure strategy has been extended to 4th grade. Since 4th grade is an important transitional point where learners move from story based reading comprehension to expository texts—(aka) content area reading, it is an important time to teach them about text structure. Finally, the design of our web-based tool, research, and outcomes were affected by on the field findings about 4th graders' ability to type. This finding required the modification of the ITSS tutor during the first month of data collection to limit the 4th graders' activities to the signaling word and main idea tasks.

The main focus of the development of the web-based ITSS was to improve expository reading comprehension. The results from the primary research question show that the students using ITSS for 30–45 min each week for approximately 6 months showed statistically significantly better performance on the researcher designed post-test measures compared to their control counterparts. These findings were obtained with the restricted version of ITSS used by the 4th graders receiving training on signaling word identification and writing main ideas. Results showed statistically significant and moderate ES for all post-test measures for which students had received instruction and practice. These included the Signaling test and main idea quality. There were statistically significant but small ES for the full recall tasks for which students did not receive any instruction or practice. A small ES was found on the standardized reading comprehension test (GSRT), consisting of mainly narrative texts with multiple-choice questions. Both the recall and GSRT findings suggest some transfer from instruction and practice on identifying signaling words about text structure and composing main ideas using the structure strategy.

Meyer et al. (2010) reported a larger effect on the standardized reading comprehension test (GSRT) with 5th-grade students with larger ES. The performance of the 4th-grade students may be explained by developmental levels of understanding and/or the limited practice they received within the modified ITSS system due to their lack of typing skills. The lack of practice may be further explained by comparing the 5th-grade results where the tasks for which students received practice (signaling, main ideas, and full recalls) all showed moderate ES. The 4th-grade students in the current study also showed moderate effects for the tasks for which they had received practice within ITSS.

Meyer et al. (2010) reported larger gains with 5th grade students using ITSS, where fifth-grade students worked on ITSS three times per week for 6 months. In contrast, students in the current study used ITSS just once a week for 30–45 min. In spite of the much shorter time spent on ITSS in the current study, ITSS classrooms performed significantly higher than control classrooms though ES were smaller. This suggests that more frequent use of the ITSS might be more beneficial.

The overall findings from this large-scale randomized controlled trial on the use of a web-based delivery tool for the structure strategy are unique and should be considered in light of many recent large-scale randomized controlled trials where findings of no significant difference were commonly reported. James-Burdumy et al. (2009) conducted 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). All four interventions showed no significant differences between the control classrooms and classrooms after a full year of intervention.

Results from this study support the efficacy of delivering the reading comprehension strategy to a large audience using web-based ITSSs. This has practical implications for schools where teachers may not be able to attend to each child in a large group setting. Schools are also investing in their technology infrastructures and tools such as ITSS can make a practical impact on the educational outcomes of students.

Findings from this research study are promising with ITSS effects being statistically and practically significant. They are particularly important in that only about 35 min a week of instruction for 6–7 month with ITSS yielded positive results for 4th-grade children. Also, notable is that the ITSS instruction helped both girls and boys write good main ideas, but the effect was stronger for boys. Boys perform less proficiently than girls on verbal tasks assessed by writing (e.g., Halpern 2006). The especially assistive aid for boys in writing main ideas may have resulted from the systematic approach to writing a good main idea taught in ITSS with a wide variety of topics of interest to boys, girls, and both boys and girls (Meyer et al. 2010). These findings combined with previous studies on the structure strategy point to an important role for the structure strategy within school language arts curricular. In light of the findings of this study with students completing the partial structure strategy training (signaling words and main idea without the full recall tasks) compared to other ITSS studies, schools implementing the ITSS system should try to use the complete system to achieve the full benefits in reading comprehension. Additionally, our in situ findings that 4th graders lacked the necessary keyboarding skills, suggests that further extensions to ITSS may be necessary to accommodate this need. A simple solution would be for students to receive keyboarding instruction in 3rd and 4th grade in preparation for using software, such as ITSS. Other examples of extensions to accommodate these students within ITSS may include voice recognition software to allow students to dictate their recall and thinking process. Another extension may be tablet PCs that allow students to write their recall using a pen-like device instead of typing on the keyboard. This would require additional software to convert the hand-writing to text.

The results from this study also are promising for technology-based learning environments. This study found statistically significant improvements in learning as measured by standardized tests as well as researcher designed measures while varying only the reading comprehension strategy and delivery medium. While this study cannot disaggregate whether the strategy or the medium played a larger part it is worthwhile to note that the structure strategy has consistently been effective in many previous studies (Meyer et al. 1980, 2002, 2010, 2011; Williams et al. 2005, 2009). Additionally, previous smaller scale studies on ITSS have also shown statistically significant improvements in grades 5 and 7 (Meyer et al. 2010). This is the first extension to fourth grade and also the largest randomized controlled trial to date on the approach delivered using web-based intelligent tutoring technologies. In contrast to many previous debates in this journal about no-significant differences in learning outcomes when technology based learning environments were introduced (Jonassen et al. 1994; Kozma 1994), this study presents evidence that carefully designed technology-based learning environments that use a well-researched underlying method, such as the structure strategy, can produce improvements in learning.

Finally, further research and analysis will be necessary to tease out the specific causes for the improvements noted. Additionally, research on motivation and self-efficacy of the learners using a web-based system is needed.

Acknowledgments The research reported here was supported by the Institute of Education Sciences, U.S. Department of Education, through Grant R305A080133 to The Pennsylvania State University Beaver. The opinions expressed are those of the authors and do not represent views of the Institute or the U.S.

Department of Education. More information on our project is available at http://itss.br.psu.edu/. The authors appreciate the contributions of the elementary students, faculty, and administration involved in this research effort. Additionally, the authors appreciate the input of other faculty, affiliates, and students at Penn State including James Spielvogel, Kathryn Shurmatz, Lori Johnson, Janessa Weaver, Melissa Ray, and Michael Cook.

References

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.

Atkinson, R. K., Mayer, R. E., & Merrill, M. M. (2005). Fostering social agency in multimedia learning: Examining the impact of an animated agent's voice. *Contemporary Educational Psychology, 30*, 117–139.

Baddeley, A. D. (1986). *Working memory*. Oxford: Oxford University Press.

Caccamise, D., Snyder L., Kintsch W., Allen C., Kintsch E., & Oliver, W. (2010). Teaching summarization via the Web. In *Annual proceedings of American Educational Research Association*.

Chall, J., Jacobs, V., & Baldwin, L. (1990). *The reading crisis: Why poor children fall behind*. Cambridge, MA: Harvard University Press.

Halpern, D. F. (2006). Assessing gender gaps in learning and academic achievement. In P. A. Alexander & P. H. Winne (Eds.), *Handbook of educational psychology* (2nd ed., pp. 635–653). Mahwah, NJ: Lawrence Erlbaum Associates.

Harp, S. F., & Mayer, R. E. (1998). How seductive details do their damage: A theory of cognitive interest in science learning. *Journal of Educational Psychology, 90*(3), 414–434.

House Commission on Rural Education (HCRE). (2004). Report to the Pennsylvania General Assembly from the Commission on the Study of Education in Rural Pennsylvania. By order of House Resolution 8, February 2003.

Jacobson, M. J., Maouri, C., Mishra, P., & Kolar, C. (1995). Learning with hypertext learning environments: Theory, design, and research. *Journal of Educational Multimedia and Hypermedia, 4*(4), 321–364.

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 June 11, 2011, from http://www.mathematica-mpr.com/publications/pdfs/education/selectsupplreading.pdf.

Jonassen, D. H., Campbell, J. P., & Davidson, M. P. (1994). Learning with media: Restructuring the debate. *Educational Technology Research and Development*, *42*(3), 31–39.

Kozma, R. B. (1994). Will media influence learning? Reframing the debate. *Educational Technology Research and Development*, *42*(2), 7–19.

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

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., & 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*(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.

Meyer, B. J. F., Young, C. J., & Bartlett, B. J. (1989). *Memory improved: Reading and memory enhancement across the life span through strategic text structures*. Hillsdale, NJ: Lawrence Erlbaum.

National Assessment of Educational Progress (NAEP). (2007). Available at http://nationsreportcard.gov/reading_2007/. Accessed 10 June 2010.

Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods* (2nd ed.). Thousand Oaks, CA: Sage.

Raudenbush, S. W., Bryk, A. S., & Congdon, R. T. (2008). *HLM 6.0 Hierarchical linear and non-linear modeling [Computer software]*. Lincolnwood, IL: Scientific Software International.

Shaywitz, B. A., Holford, T. R., Holahan, J. M., Fletcher, J. M., Stuebing, K. K., Francis, D. J., et al. (1995). A Matthew effect for IQ but not for reading: Results from a longitudinal study. *Reading Research Quarterly, 30*, 894–906.

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

Sweller, J. (2005). Implications of cognitive load theory for multimedia learning. In R. Mayer (Ed.), *Cambridge handbook of multimedia learning* (pp. 19–30). New York: Cambridge University Press.

Theodorou, E. (2005). *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, The Pennsylvania State University.

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

Williams, J. P., Hall, K. M., Lauer, K. D., Stafford, K. B., DeSisto, L. A., & DeCani, J. S. (2005). Expository text comprehension in the primary grade classroom. *Journal of Educational Psychology, 97*(4), 538–550.

Williams, J. P., Stafford, K. B., Lauer, K. D., Hall, K. M., & Pollini, S. (2009). Embedding reading comprehension training in content-area instruction. *Journal of Educational Psychology, 101*, 1–20.

Dr. Kausalai Kay Wijekumar is Associate Professor of Information Sciences and Technology at The Pennsylvania State University Beaver. Her research focuses on *learning with* technologies and the *effects of* technologies on learners. She has led large scale randomized controlled trials for the U.S. Department of Education, Institute of Education Sciences (Improving reading comprehension in Grade 4–8 using web-based intelligent tutoring systems) and the Regional Educational Laboratory Mid-Atlantic (Odyssey Math – 4th grade). She is an elected member of the Board of School Directors in her local district and is a passionate advocate for effective learning tools for children.

Dr. Bonnie J. F. Meyer Professor of Educational Psychology at The Pennsylvania State University, studies reading comprehension, text structure, discourse processing, decision making, and the structure strategy. She has over 50 publications about the structure strategy and its effects on understanding and applications for reading, writing, and everyday problem solving. Her current research, funded by the Institute for Education Sciences of the U.S. Department of Education, is focused on structure strategy instruction for students in Grades 4-8. She has served on editorial boards for *Discourse Processes*, the *Journal of Educational Psychology*, the *Reading Research Quarterly*, *Cognition and Instruction*, the *Educational Psychologist*, and the *Journal of Literacy Research* and is a Fellow in the American Psychological Association, American Educational Research Association, and American Psychological Society.

Dr. Puiwa Lei is Associate Professor of Educational Psychology at The Pennsylvania State University. She teaches courses in advanced statistics and measurement theories. Her research interests are in methodological issues of multivariate statistical analysis techniques. Dr. Lei has conducted data analyses for large scale randomized controlled trials for the U.S. Department of Education Regional Educational Laboratories and Institute of Education Sciences.