Computers & Education 68 (2013) 366–379

Contents lists available at SciVerse ScienceDirect

Computers & Education

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

High-fidelity implementation of web-based intelligent tutoring system improves fourth and fifth graders content area reading comprehension

Kausalai (Kay) Wijekumar [a,*], Bonnie J.F. Meyer [b], Puiwa Lei [b]

[a] College of Information Sciences and Technology, The Pennsylvania State University Beaver, 212 Ross Administration Building, Monaca, PA 15061, USA
[b] Department of Educational and School Psychology and Special Education, The Pennsylvania State University, USA

A R T I C L E I N F O

Article history:
Received 20 February 2013
Received in revised form
22 April 2013
Accepted 23 May 2013

Keywords:
Reading comprehension instruction
Structure strategy
Fourth- and fifth-grade reading
Web-based tutoring
Fidelity of implementation

A B S T R A C T

Technologies and their effectiveness are impacted by how well they are implemented. A large scale randomized controlled trial was conducted to study the efficacy of a web-based intelligent tutoring system to deliver the structure strategy to improve content area reading comprehension. We present our theory of change focusing on the theoretical framework: structure strategy, delivery approach of web-based intelligent tutoring systems, and contextual conditions for successful adoption of the tool with fidelity. Results from the optimal implementation schools show statistically significantly better performance by ITSS classrooms compared to their control counterparts with moderate to large effect sizes. Conditions for implementing technology-based interventions with fidelity in schools are discussed.

© 2013 Elsevier Ltd. All rights reserved.

1. Introduction

The hype cycle for technologies (Gartner, 2012) characterizes how emerging technologies are expected to produce amazing results moving through a period of disillusionment to an eventual role in the domain. Additionally, Gartner's research shows that over 90% of computer tools do not achieve their intended outcomes. Computer aided learning environments are no exception to this rule but appear not to settle into any pattern due to the numerous complex factors contributing to their success and failure. Even though scholarly journals are full of technology-based solutions that appear to produce large effect sizes in small studies, many do not reach their forecasted potential in large scale randomized controlled trials (e.g., Drummond et al., 2011; Wijekumar, Hitchcock, Lei, Turner, & Peck, 2009).

Interventions to improve reading comprehension using teacher-led and computer-based delivery media also have not achieved their purported goals in large-scale studies (Burdumy et al., 2009). Further, forty plus years of reading interventions and curricula (including computer-based tools) have only shifted the reading comprehension scores on the national assessment of educational progress by a few points since the first tests in 1969. The 4th grade average reading score was 217 in 1992 and 221 in 2011 (NAEP, 2011). The 8th grade average reading score was 260 in 1992 and 265 in 2011 (NAEP, 2011). There are many reasons why the current instructional approaches and computer-based tools have not succeeded in achieving better outcomes in national tests and large scale studies. These reasons include lack of (a) sound theoretical basis for the intervention, (b) instructional methods that support the development of skills, (c) multi-pronged solutions that cater to the diverse needs of learners, (d) effective use of media for delivery of the intervention, and (e) fidelity in implementing the curriculum and/or software. The focus of this study was fidelity in implementing the curriculum and/or software.

Our focus is on elementary level reading comprehension at grades 4 and 5 as well high-fidelity web-based delivery of a reading comprehension approach – the structure strategy that has shown promise in improving reading comprehension for over 40 years (e.g., Armbruster, Anderson, & Ostertag, 1987; Bartlett, 1978; Carrell, 1985; Cook & Mayer, 1988; Englert & Hiebert, 1984; Gordon, 1990; Meyer, Brandt, & Bluth 1980; Meyer et al. 2002; Meyer, Wijekumar, & Lin 2011; Meyer et al., 2010; Meyer, Young, & Bartlett, 1989; Paris, Cross,

* Corresponding author.
E-mail addresses: Kxw190@psu.edu, kaywije@verizon.net (K.(Kay) Wijekumar), Bjm8@psu.edu (B.J.F. Meyer), Pul3@psu.edu (P. Lei).

0360-1315/$ – see front matter © 2013 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.compedu.2013.05.021

& Lipson, 1984; Polley, 1994; Raphael & Kirschner, 1985; Richgels, McGee, Lomax, & Sheard, 1987; Samuels et al., 1988; Weisberg & Balajthy, 1989; Wijekumar, Meyer, & Lei, 2012; Williams et al., 2005; Williams, Stafford, Lauer, Hall, & Pollini, 2009).

The smaller research studies on the structure strategy presented by trained teachers or the web-based intelligent tutoring system for the structure strategy (ITSS) have shown strong results with moderate to large effect sizes (Meyer et al., 2010; Meyer et al., 2011). Meyer et al. (2010) showed that ITSS with elaborated feedback showed higher gains on a standardized test of reading comprehension – *Gray Silent Reading Test* (GSRT; Wiederholt & Blalock, 2000) ($d = .55$) than student receiving simple feedback ($d = .15$). The comparison text structure competence in the study showed an increase of $d = .61$ from pre- to post-test. In the 2010 study, students used ITSS for 2 class periods a week with approximately 90 min of instructional time. During these 90 min students received modeling, practice, assessment, and feedback from the animated pedagogical agent in ITSS named Intelligent Tutor (I.T.). Similarly, Meyer et al. (2011) used an individualized version of ITSS and found large gains. In contrast, recent large-scale studies on ITSS resulted in modest effect sizes of .19 and .20 on the GSRT and .18 and .26 on comparison competency, for 4th and 5th grades, respectively (Wijekumar et al., 2012, 2013). The purpose of the current research was to examine whether fidelity of treatment delivery affects the learning outcomes and strength of findings.

We present a theory of change (shown in Fig. 1) suggesting that a strong theoretical foundation provided by the structure strategy, optimal instructional strategies to present the learning, a carefully designed web-based intelligent tutoring system as a delivery medium, and implementation with fidelity by the teachers and school personnel can achieve expected results on reading comprehension. Finally, we present results from a randomized control study with standardized and researcher-designed outcome measures.

2. Framework for improving reading comprehension

As seen in Fig. 1 the structure strategy provides the foundation for the ITSS intervention. Key instructional strategies for ITSS involves the modeling of expert use, scaffolded practice, assessment, and individualized and elaborated feedback.

2.1. Sound theoretical framework

The theoretical framework for the structure strategy focuses on the selection and encoding of text to create tightly organized strategic memory representations that foster efficiencies in storage and retrieval. The process of selecting important ideas from the text is assisted by finding the text structure organizing the passage. Students learn to classify the organizing text structure as one of the five identified by Meyer (1975; Comparison, Problem and Solution, Cause and Effect, Sequence, and Description) through first finding signaling words.

Using each text structure's organizing framework for main ideas (e.g., comparison framework: ____ and ____ were compared on ____, ____, and ___), students create mental representations as well as written summaries (main idea) that start the encoding process at the highest level (Meyer & Poon, 2001). Once this top level encoding is established and connected to existing memories, students further populate the hierarchical tree with details from the passage. Thereby strategically encoding the passage within memory in an efficient manner and making it easy to access and utilize when needed for recalling, writing, applying, inferring, and problem solving.

Three important features distinguish the structure strategy from other practices. First, the structure strategy applies to all content domains. A comparison structure can be used to showcase differences and similarities among invertebrates (biology domain) as well as the northern and southern States' responses to the civil war (history domain). Further, the many representations of the structure strategy's main idea (i.e., matrix, paragraph, concept map, Venn diagram) allow students to organize their memories as well as their laboratory reports, class notes, and observations. It enhances his/her argumentation skills by focusing attention to problems and solutions, causes for the problems, and comparisons of solution alternatives.

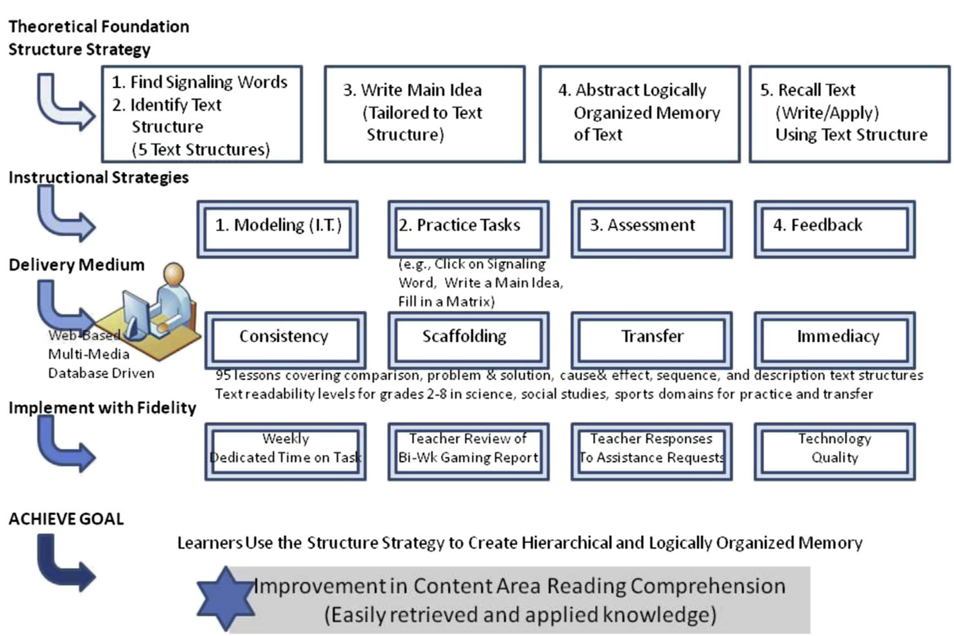

Fig. 1. Theory of change for web-based structure strategy delivery and reading comprehension outcomes.

Second, the structure strategy can help during multiple steps of processing during reading, comprehending, and applying the knowledge. During *initial reading* the structure strategy is designed to focus students' attention to important ideas in the text as well as *encoding and creating effective memory structures*. The method can assist students when they perform reading comprehension activities such as summarizing, underlining, and building background knowledge. Later in the reading process it can further assist in *comprehension monitoring* tasks that are frequently a challenge to many students. By traversing his or her mental representation of the text and using the text structure's organization, students learn how to fill in gaps in their knowledge and also check their understanding.

Because of the flexibility of this approach, the structure strategy enhances most comprehension curricula across the country. For example, if teachers use highlighting important ideas or summarizing as comprehension development activities, the structure strategy can help identify what to highlight. In summarizing, the structure strategy not only helps in finding the important ideas to include but provides a framework for connecting the ideas together. When students write or when they try to apply their knowledge in problem solving, the hierarchical memory structure is efficient and effective. In contrast to loosely organized memory structures (frequently used by poor readers, Meyer et al., 1980) that have no logical glue binding the ideas together, the structure strategy allows students to a) remember key top-level ideas (increasing efficiency) and b) further associate relevant subordinate ideas when needed by traversing the hierarchical conceptual tree (making it associative and effective). These types of efficiencies are frequently cited in expertise research for reasons why experts are better problem solvers than novices (Chi, Feltovich, & Glaser, 1981).

Third, the structure strategy has been understood and applied by students as young as second grade (Williams et al., 2009) and older adults with low literacy skills (Meyer et al., 2001). When adults teach children about reading comprehension and ask them to perform tasks like finding important ideas or generating questions about the text, students gravitate to isolated details in the text until they become quite familiar with the text. The structure strategy can guide the learner in developing expertise in reading comprehension at a faster rate than repeatedly practicing complex tasks with little to guide the student other than a teacher's general request for "what is important in the text".

Research on the structure strategy has been consistent in improving comprehension outcomes whether with primary grade students, undergraduates, or older adults. These findings have been replicated over 40 years (e.g., Meyer, 1975; Wijekumar et al., 2012). While many variations to the approach and delivery have shown improvements, Meyer's approach has been thorough and broad (e.g., grade levels, literacy levels, domains, age ranges, and prior knowledge).

ITSS was designed to widely disseminate these research findings and uses Meyer's approach to text structure. Fig. 2 presents a high-level view of the system's features. Our team has reviewed over 10 reading curricula used in classrooms participating in the current study (e.g., Scott Foresman Reading Street, Harcourt Trophies) and found little to no formal text structure instruction. Also, the researchers conducted professional development for teachers participating in the current study. A common sentiment heard at many of the professional development sessions and school presentations was that teachers were aware of problem and solution text structures and knew of "signal words." Most did not complete the cycle specified in the structure strategy (Meyer et al., 1989) – find the signaling words, classify the text structure, write/abstract a main idea specific to the text structure, use both the main idea and signaling words to build a coherent memory structure and written representation of the text. This cycle for structure strategy instruction yields useful procedural knowledge rather than just declarative knowledge about definitions for text structures and signal words. The structure strategy as presented by the web-based ITSS is broad – it covers many domains, showcases application of nested text structures, provides scaffolded practice with grade appropriate reading passages, provides transfer tasks to real-life texts, assesses student responses immediately, and gives student advanced, elaborated feedback on his/her performance on each task (Meyer et al., 2010; Wijekumar et al., 2012).

2.2. Instructional methods that support the development of skills

Students interact with ITSS and learn how to use the structure strategy to comprehend content area texts. The enabling learning goals include finding signaling words in a passage (see Fig. 3), learning to identify which text structure uses those signaling words, using a main idea pattern to construct a summary and strategic representation for the text (see Fig. 4), writing a recall of the text without access to the

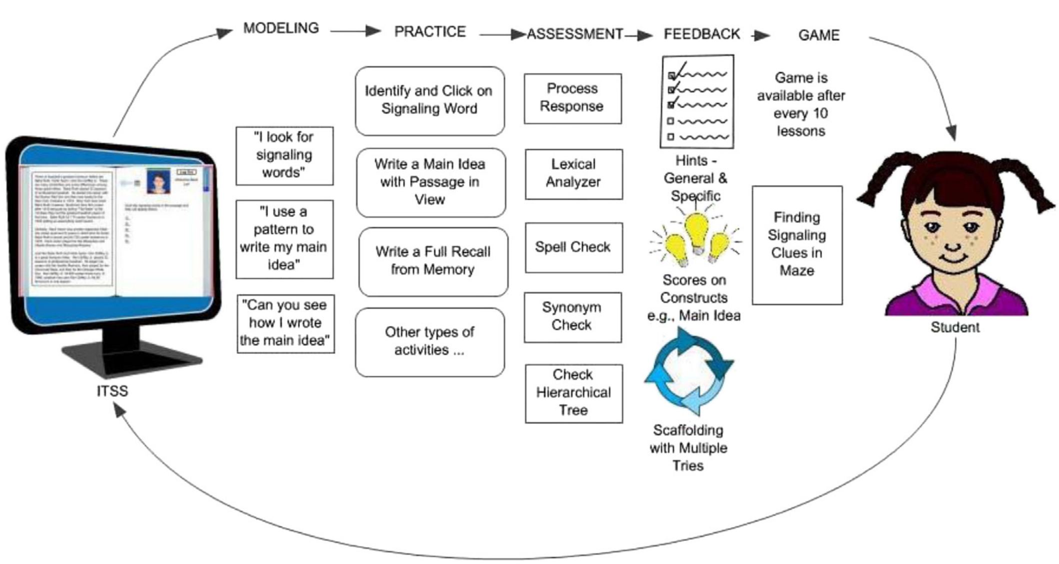

Fig. 2. High-level view of ITSS system.

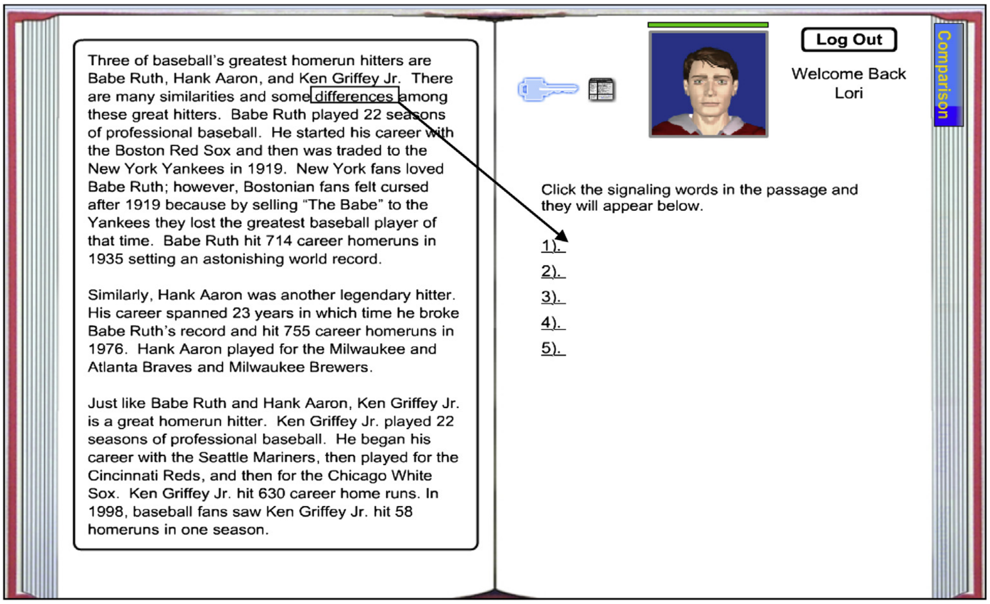

Fig. 3. ITSS page showing signaling word task.

passage (from memory), and reviewing/monitoring understanding by checking his or her organizing structure and completeness of recall. Students learn to use "keys" that serve as job-aids with signaling words and text structures shown in Fig. 5.

Students practice these skills by reading passages in many different domains and also with texts of varying complexity (i.e., readability levels, number of embedded text structures). The text structures are presented in an order – comparison, problem and solution, cause and effect, sequence, and description based on previous research (Meyer et al., 2002). Students receive practice in applying their skills within each text structure before proceeding to the next text structure. The first 10–12 lessons are dedicated to showing the use of the comparison text structure with topics, such as elephants, favorite U.S. Presidents, Olympic athletes, and whales. As students proceed to learning about multiple text structures, they learn how to combine text structures. After approximately 20 lessons students would have completed both the comparison and problem and solution text structures and are shown how to combine those text structures in comprehension and writing.

During each lesson students learn how to find signaling words and classify the text structures, write main ideas with main idea patterns, write recalls, and monitor their understanding of the text. Students practice their skills in answering questions and receiving customized feedback to scaffold their learning. Even though these instructional strategies can be used in teacher-led classroom settings, ITSS was designed to broaden the reach of the approach with consistency and depth. Training a large number of teachers to deliver the structure strategy with consistency would be a difficult task. Further, ITSS provides a variety of examples from many difference domains and individual attention to the student. The consistency, numerous examples, and individual attention are difficult to accomplish widely in a teacher-led classroom setting.

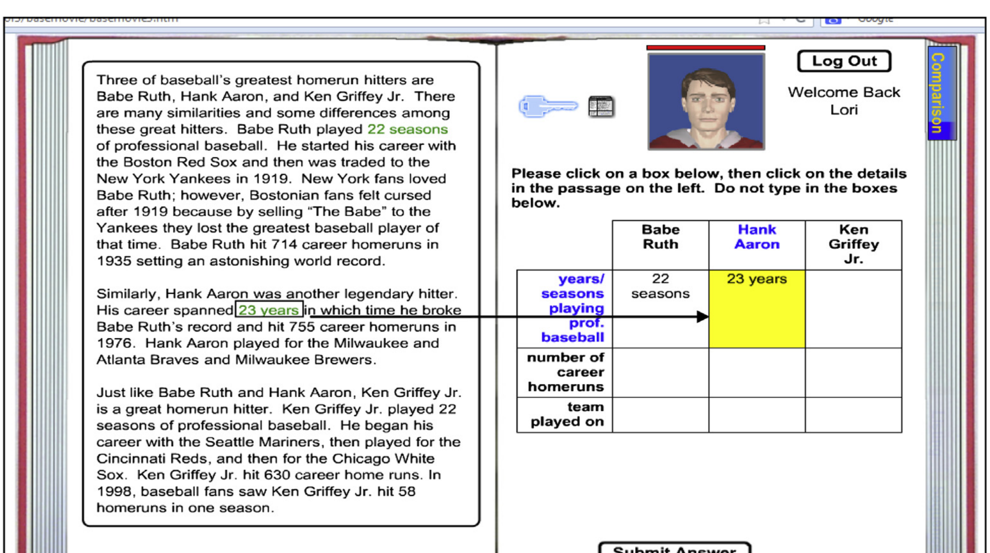

Fig. 4. ITSS page showing main idea matrix task.

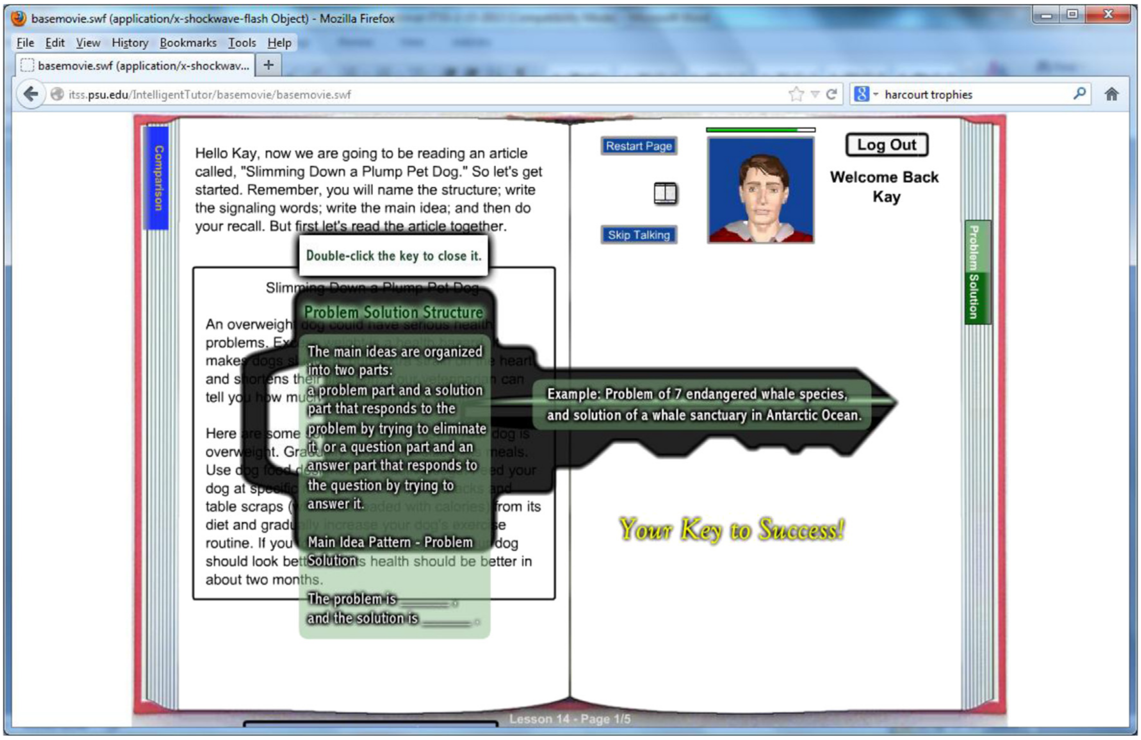

Fig. 5. Problem and solution text structure key (front).

2.3. Effective use of media for delivery of intervention

The plethora of available multi-media tools requires designers to carefully juxtapose the learning goals with available media to make sure that the learning goals drive the multi-media choices. In designing ITSS, the focus has always been on applying the intelligent tutoring framework, multi-media, and web-based interface to help students learn how to use the structure strategy. Our design provides a book-like affordance for the learner as shown in Fig. 3. Further, narrations with human voice accompany every page allowing learners at differing levels of skills to read the text as well instructions (Atkinson, Mayer, & Merrill, 2005; Mayer, 2009). This allows students with lower verbal skills to read along with the narration. It also ensures that all students will have to attend to what is being presented on the screen before they move on to the next page. When students have the ability to press the "next" button on a lesson, some choose to ignore the contents of the page and keep pressing "next" till the questions appear. The ITSS screen's navigation buttons do not appear on the screen until I.T. has completed his narration of the page contents.

Interactions between the Intelligent Tutor (I.T.) and the student are initiated by I.T. and proceed with frequent modeling in early lessons for each text structure through Flash animations and highlighting important ideas through colors and words blinking on screen (shown in Fig. 2). I.T. then asks students to find signaling words, write main ideas, fill in a matrix, write a full recall, and reviews his/her work. Once students have completed each task, I.T. assesses his or her response using simple keyword matches for simple responses. When students write main ideas or full recall responses, the scoring system uses a series of steps including parsing, spelling checks, checks for synonyms, and also comparing the response to the hierarchical conceptual tree for a passage. Scores are generated on multiple constructs such as signaling words, main idea, and details (subordinate ideas).

The multi-media use is carefully designed to eliminate irrelevant images or animations that may cause distractions to students. Mayer (2009) describes these elements of multi-media learning environments as seductive details that draw the students' attention away from the learning goals. Instead, histograms of stars show the learner how well they performed on tasks and tabs on the book fill up with colors as they make progress showing the student his/her relative position within the lessons.

Games were designed to fit into the learning scheme and motivate the students (see Fig. 2 for placement of games). After approximately 10 lessons students play a game of finding signaling words with a dinosaur. Trophies also pop-up when students complete tasks on the first try and receive a high score.

Feedback is also narrated by I.T. with advanced hints, pop-ups that assist in improving the students' performance, and scaffolding. As students work on practice tasks the number of tries for each task is varied. This helps reduce gaming because students cannot expect feedback to appear after a pre-determined number of tries.

2.4. Fidelity in implementing the software

The thoughtful design and development of any intervention is only one part of achieving the intended learning outcomes. Fidelity of implementation is an essential part of delivering the intervention to learners as the designers and developers intended. One of the appeals of computer-based tools is that consistency can be achieved in delivering the intervention. Variations in instructions delivered by the teacher, such as examples, practice tasks, and scaffolding, can be placed on an even playing field through web-based delivery. A lack of examples, practice tasks, and scaffolding may only be one part of the problem. A bigger problem may be that teachers may not be teaching reading comprehension strategies in their classrooms as noted by Pressley, Wharton-McDonald, Mistretta-Hampton, and Echevarria (1998).

Delivering the intervention using technology also allows each child to learn at his or her own pace with one on one tutoring that cannot be achieved in a group environment. Finally, students who have difficulty reading may feel uncomfortable in a group setting if they are called on by a teacher to summarize or answer questions about the passage. In one-on-one tutoring with the computer, the students' lack of knowledge or skill is not highlighted to the rest of their peers.

Unfortunately, all the efficiencies that can be achieved through web-based delivery also come at a cost. In the case of the poor learning outcomes reported in large scale randomized controlled trials there appear to be factors beyond a good tool and available technologies that allow the system to achieve the intended outcomes. One area that is frequently cited as the reason for the poor learning outcomes in large scale studies is fidelity of implementation (Wijekumar et al., 2009).

Fidelity of implementation can be measured using quantitative metrics such as time on task, teacher responses to questions from students, and teacher reviews and acting on the bi-weekly gaming reports. Unfortunately, there are elements of fidelity, such as teacher knowledge and interest in the activity, that are not measured as easily. Our approach for this study was to concentrate on the measurable constructs of fidelity.

During the course of the data collection for this research study, teachers were mailed bi-weekly reports on student progress, performance on tasks, and gaming. The reports highlighted exceptional performance by students as well as any gaming patterns where students may type in nonsense answers or blanks. Teachers were encouraged via emails to follow-up and address any concerns the students had or remind students to pay attention to the lessons. The research team monitored the students' changes in gaming after reports were mailed to the teachers. Some teachers contacted the research team and followed up on the progress of the students. Teachers who exhibited this pattern were considered to have met the first criterion for implementing the software with fidelity. For example, such teachers asked follow-up questions to the research team (e.g., can you tell me how many times student ABC1000 has gamed the system?") and also motivated the students by encouraging them with information from the reports (e.g., "The entire class has done really well in completing the lesson 10").

In addition to these reports, the trained research personnel conducted two fidelity observations in all classrooms during the middle of the academic year. During these observations, the teachers' ability to monitor and answer questions from students was the second criterion for implementing the software with fidelity. Teachers who implemented the software with fidelity understood how the software worked, were not affected by any hardware or software malfunction (e.g., file loading slowly), monitored the students by walking around the classroom and asked students follow-up questions about the lesson they were completing, and were able to help the student if they were experiencing any difficulties.

The third criterion for inclusion in the high-fidelity implementation group was the consistent time spent on the ITSS lessons. The classrooms needed to have used ITSS each week (except holiday breaks) for at least 30 min each week over approximately six months. The classrooms that implemented ITSS with fidelity had an average time on ITSS of 35 min each week compared to the other classrooms that provided only 20 min or less of ITSS each week.

In summary, fidelity of implementation was measured by the teachers' attention to gaming and progress reports, observations that showed the teachers' ability to monitor students working on ITSS, and weekly sustained use of the software. All three criteria had to be met to be included in this analysis.

3. Research questions

The purpose of this study was to examine whether ITSS delivered with fidelity can improve reading comprehension of 4th and 5th grade students as measured by standardized tests and researcher designed measures. The research questions were:

Do 4th and 5th grade classrooms using the ITSS delivery of the structure strategy with fidelity as a partial substitute for the standard language arts curriculum outperform control classrooms on a standardized measure of reading comprehension?

Do 4th and 5th grade classrooms using the ITSS delivery of the structure strategy with fidelity as a partial substitute for the standard language arts curriculum outperform control classrooms on researcher designed measures of reading comprehension?

4. Method

4.1. Research design

A multi-site cluster randomized trial was conducted with rural and suburban 4th and 5th grade classrooms. Fig. 6 illustrates the design. Experimental and control groups actually had the same curriculum (except for a class period each week) because classrooms were randomly assigned to ITSS or control groups within schools. One of the regular language arts class was substituted by ITSS instruction each week for the experimental group while the control group continued with their regular language arts class. The diagram shows the consistency in curriculum and instructional time within the school. Schools agreed to use the software as a partial substitute to the Language Arts classrooms for 30–45 min each week.

A volunteer sample of 128 teachers and their 5th grade classrooms were randomly assigned to the ITSS and control groups within schools. Similarly, 131 teachers and their 4th grade classrooms were also randomly assigned to the ITSS or control groups within the schools. Schools signed a memorandum of understanding agreeing to participate in the research study.

The ITSS classrooms selected for this analysis were those that implemented the software with fidelity as described earlier. A total of 5 schools and 24 classrooms were used in this analysis. The analysis used both the ITSS and control classrooms within the same school to maintain the advantages of the within school comparisons.

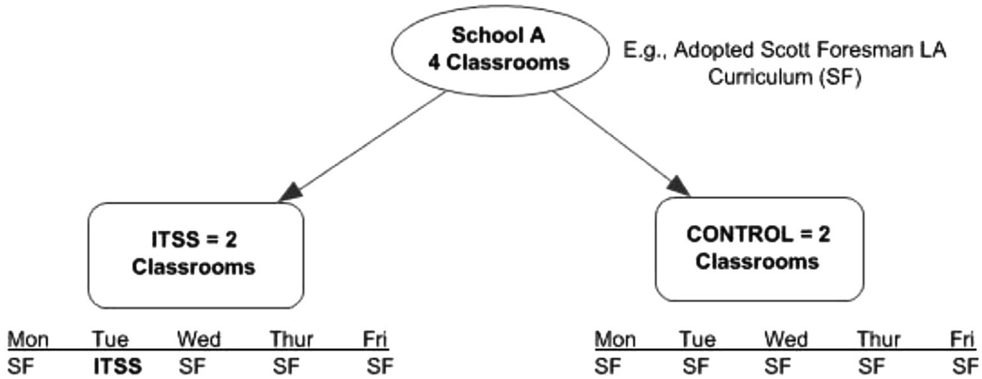

Fig. 6. Example of within school random assignment of classrooms maintaining instructional time, and curriculum.

4.2. Participants

Participants were recruited through direct mailing to Superintendents, telephone calls, and through presentations at regional locations e.g., intermediate unit locations. The research team was assisted in this recruitment by member of the Laboratory Extension Specialists at The Pennsylvania State University. In order to participate in the research, schools had to have the computer laboratories and sufficient bandwidth to support the implementation of the ITSS software for entire classroom of students.

Site visits were conducted prior to the signing of a memorandum of understanding to verify the availability of computers and test the bandwidth with 20–25 concurrent users. School technology personnel also were present during this site visit to check on the correct versions of the Flash player, browser settings, and white-listing of the ITSS web site on the schools' firewalls.

At the conclusion of the recruitment effort a total of 45 schools signed the agreement to participate in the research study. This analysis uses the five schools that implemented the ITSS intervention with high fidelity. These schools had 64% of 4th grade students and 58% of 5th grade students reading at advanced or proficient levels (on the State assessments). The schools had 49.8% economically disadvantaged and 29% minority students. None of the schools had any English language learners.

All 4th and 5th grade teachers in the participating schools were invited to participate, and none declined. All the participating teachers (classrooms) were randomly assigned to the ITSS and control conditions after students had been assigned to teachers in the schools. Teachers completed their consent forms at the professional development sessions or during the site visits by the study team. Each school mailed parental consent forms to all students at the 4th and 5th grade level prior to notification of random assignment.

4.3. Procedure

All students who had parental permission completed standardized and researchers designed measures of reading comprehension at the beginning of the academic year. Students completed the *Gray Silent Reading Test* (Wiederholt & Blalock, 2000) and researcher-designed measures of reading comprehension prior to the start of the intervention. The tests were administered by the research team with the support of the teachers in large group settings in a quiet area.

Students in the intervention condition from the high fidelity schools used ITSS at least once a week for 30–45 min each week over a 6–7 month period during the 2009–10 academic year. Each teacher kept the student login sheets in his or her classroom and handed out the sheet to each student at the beginning of the class period. Students put on their earphones, logged into the computer and clicked on the link to ITSS on the web browser to begin their session. I.T. initiated the conversation with the student by starting the student where they stopped during the previous ITSS session.

Students completed the post-test measures of reading comprehension – GSRT and researcher designed, at the end of the school year. The tests were administered by the research team with the support of the teachers in large group settings similar to the pre-test.

4.4. Materials

The outcome measures used at pre- and post-test included two forms of a standardized test of reading comprehension and two forms of a researcher designed measure.

The standardized tests of reading comprehension were forms A and B of the *Gray Silent Reading Test* (Wiederholt & Blalock, 2000). Form B was administered at pretest and Form A was administered at post-test. The pretest score on the GSRT was used as a covariate for data analyses used to examine the effects of ITSS instruction on the dependent measures that focus on reading comprehension. Cronbach's alpha for both forms of the GSRT was alpha = .88. The GSRT has fifteen narrative texts with each text followed by five multiple-choice questions.

Experimenter-designed measures of reading comprehension used two equivalent test forms created by Meyer et al., 2010. Form 1 was administered at pretest before the children started ITSS and the Form 2 at post-test after completing the ITSS intervention (approximately 6 months later). A summary of the form contents as well as scoring guideline examples are presented in Table 1.

4.5. Scoring

Trained graduate students who scored the main idea and recall tasks were blind to the experimental condition of the participants. Meyer's scoring procedure (1975, 1985) using a propositional analysis of ideas in the text and hierarchical relationships were used in the

Table 1
Researcher designed outcome measures, scoring approaches with examples, and inter-rater agreement.

Construct measured	Scoring approach and examples of scoring guidelines	Inter-rater agreement
Problem and Solution Source: 2 Equivalent Problem and Solution Passages each with 98 words and 72 idea units		
Problem and Solution Competency and Full Recall (without the passage in view) At least two trained raters and computer scoring		
Problem and Solution Competency	Score 1–8 1 = no problem, no solution, and no cause 2 = signaled cause but no problem and no solution 3 = one part of the problem and solution 4 = problem and cause but no solution or incorrect solution 5 = problem and solution (correct content of problem) and correct content of solution 6 = problem, solution & cause of the problem mention only in the solution part. 7 = similar to 5 but additionally presented the cause of the problem when discussing the problem 8 = problem, solution, and cause in the problem and cause eliminated in the solution part	89%
Problem and Solution Full Recall	Total number of ideas recalled.	98%
Comparison Text Structure Source: 2 Equivalent Comparison Passages each with 128 words and 96 idea units (1) Signaling word scores based on fill in the blanks – Cloze Task, (2) Main idea score - Write a main idea for the passage (with passage in view) (3) Comparison competence and full recall score -Write a recall (*without* passage in view) At least two trained raters and computer scoring		
Signaling Word Score (Cloze Task) 7 points max for each of the 4 fill-in-the-blank words	Computer and at least 2 trained raters scoring 1 = any word 2 = words that show understanding of two animals being compared but not signaling words (e.g., "joining") 3 = signaling words, but not for the comparison structure (e.g., "solution") 4 = comparison signaling words with different intent (e.g., "smaller than" when larger than fit the context) 5 = similar signaling word (e.g., "also like") 6 = misspelled or parts of signaling words (e.g., "same") 7 = exact signaling words (e.g., "same as")	98%
Comparison Main Idea Quality	Score 1–6 (At least two trained raters scoring) 1 = no mention of two ideas compared and no mention of what attributes they were compared on 5 = correct identification of the entities compared and at least one attribute on which they were contrasted. 6 = criteria for 5 (above) but with at least 2 attributes/issues and one of the issues was a super-ordinate issue constructed from the text	93%
Comparison Structure Competency	Score 1–8 similar to those used in problem and solution scoring but using the comparison structure – two entities compared on three attributes with accurate attribution of facts to the correct entity. For example, for a maximum score of 8 = two specific animals compared on three or more parallel, accurate, and specific traits correctly attributed to each animal with no incorrect issues	89%
Comparison Structure Full Recall	Total number of ideas recalled from the text	99%

training. Complete details of the scoring approach are presented in Meyer (1985) and Meyer et al. (2010). A summary of the constructs measured, types of scores awarded to each response, and reliability are presented in Table 1.

4.6. Data analysis

To determine if there were differences between ITSS and control classrooms with respect to reading performance for both 4th and 5th grades, a two-level hierarchical linear model (HLM: Raudenbush & Bryk, 2002) was run for each of the outcome measures (GSRT and researcher-designed measures of reading comprehension) by grade using the HLM7 software (Raudenbush, Bryk, & Congdon, 2008). Listwise deletion was used to handle missing data at the time of analysis for each model to maximize the use of available data. In each of the HLM models, student-level predictors included gender (1 = female, 0 = male, grand-mean centered), Gray pretest scores (group-mean centered), and experimenter-designed reading comprehension pretest scores (comparison recall competency ratings for all outcome measures except for signaling and comparison main idea quality ratings for which his/her pretest scores were used instead; group-mean centered). Classroom-level predictors included class-mean Gray pretest scores (grand-mean centered), the corresponding class-mean experimenter-designed reading comprehension pretest scores (grand-mean centered), and ITSS (effect coded: $1/2$ = ITSS experimental group, $-1/2$ = control group). The coefficient for ITSS was of main interest because it represented the adjusted difference between ITSS and control groups in the outcome scores controlling for reading comprehension pretest scores and students' gender. Effect size (ES) for ITSS was computed by dividing this adjusted difference (i.e., estimated coefficient for ITSS from the HLM model) by the student-level pooled standard deviation on the pretest, if available, or the student-level pooled standard deviation on outcome measure (when pretest was not available).

5. Findings

There was no statistically significant difference between ITSS and control groups on the pretests at the random assignment classroom level ($t_{21} = |.31|$, $p = .76$ for the GSRT; $t_{21} = |.36|$, $p = .72$ for the Signaling test; $t_{19} = |.16|$, $p = .88$ for Main Idea Quality; $t_{21} = |.75|$, $p = .46$ for Comparison Competency). This indicated that the ITSS and control classrooms were comparable in their reading level before the implementation of the experiment.

Simple descriptive statistics by treatment condition, statistical test results of treatment effect from HLM analyses, and effect sizes on GSRT, comparison, and problem and solution post-test scores are presented in Tables 4–6, respectively. Results are discussed by research question and presented next.

5.1. Answers to research question

To address the question of whether grade 4 and 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 (Table 1), we examined results displayed in Table 2 (4th grade HLM) and Table 3 (5th grade HLM). The 4th grade students in the ITSS classrooms on average scored 3.79 points (or .36 standard deviations) higher on the GSRT adjusted post-test scores (Table 7) than their control classroom counterparts. The 5th grade students in ITSS classrooms on average scored 2.10 points (or .19 standard deviations) higher on GSRT adjusted post-test scores (Table 7).

Student outcomes were also measured based on researcher-designed measures – *comparison* signaling words, main idea quality, and competence (Tables 5 and 7). *Problem and solution* total recall and competence were also used to analyze outcomes (Tables 6 and 7). On the *comparison structure* outcomes at the 4th grade, signaling score difference was 4.48 (or .71 standard deviations, $p < .001$), main idea quality difference was 1.26 (or .83 standard deviations, $p < .001$), total recall difference was 3.62 (or .26 standard deviations, NS), and comparison competence difference was .71 (or .31 standard deviations, $p < .05$) holding reading pretest scores and gender constant (Table 7). On the *problem and solution structure* outcomes at the 4th grade, total recall difference was 3.11 (or .31 standard deviations, $p < .05$) and competence difference was .62 (or .29 standard deviations, $p < .05$) (Table 7).

On the *comparison structure* outcomes at the 5th grade, signaling score difference was 3.12 (or .45 standard deviations, $p < .05$), main idea quality difference was 1.14 (or .74 standard deviations, $p < .001$), total recall difference was 9.59 (or .54 standard deviations, $p < .05$), and comparison competence difference was 1.19 (or .50 standard deviations, $p < .05$) holding reading pretest scores, and gender constant. On the problem and solution structure outcomes at the 5th grade, total recall difference was 4.88 (or .39 standard deviations, $p < .05$) and competence difference was .80 (or .33 standard deviations, $p < .05$) (Table 7).

The effect size of .36 at 4th grade and .19 at 5th grade on the standardized GSRT test were considered small while the effect size of .83 (4th grade) and .74 (5th grade) on the comparison main idea quality measure were considered large. The 5th grade comparison measures of signaling, total recall, and competence were medium (.45–.54). The 4th grade signaling test was large at.71 and total recall and competence were small (.26 and .31, respectively). Problem and solution total recall and competence were small at both grade levels (ranged from .29 to .39).

6. Discussion

In this randomized controlled study comparing the web-based delivery of the structure strategy (ITSS) and control classrooms with a pre- and post-test data collection, we have analyzed the 4th and 5th grade classrooms that implemented the ITSS intervention with high

Table 2
Grade 4 HLM model estimates.

Fixed effects	Gray posttest	Signaling posttest	Comparison main idea quality	Comparison total recall	Comparison recall competency rating	Problem & solution total recall	Problem & solution recall competency rating
Fixed effects							
Intercept	27.88*** (.62)	15.03*** (.37)	2.84*** (.10)	19.98*** (.89)	3.59*** (.14)	14.47*** (.60)	3.06*** (.12)
Female	−1.23 (1.36)	1.00 (.53)	.10 (.14)	2.37* (1.11)	.16 (.20)	2.24* (1.03)	.35 (.22)
Gray pretest	.50*** (.06)	.22*** (.04)	.04*** (.01)	.39*** (.07)	.06*** (.01)	.32*** (.05)	.07*** (.01)
Comparison recall competency pretest	.93** (.28)	–	–	1.75** (.48)	.19** (.07)	1.20*** (.25)	.26*** (.05)
Signaling pretest	–	.27** (.08)	–	–	–	–	–
Comparison	–	–	.15* (.06)	–	–	–	–
Class average gray pretest	.76** (.22)	.22 (.12)	.10** (.03)	.98*** (.20)	.16*** (.04)	.65* (.24)	.15* (.05)
Class average comparison recall competency pretest	.20 (1.42)	–	–	3.04* (1.36)	.31 (.19)	1.41 (.70)	.17 (.14)
Class average signaling pretest	–	.53* (.23)	–	–	–	–	–
Class average comparison	–	–	.55* (.20)	–	–	–	–
ITSS	3.79** (1.19)	4.48*** (.75)	1.26*** (.19)	3.62 (1.90)	.71* (.30)	3.11* (1.20)	.62* (.26)
Random effects (Variances of)							
Classrooms	4.63*	1.45*	.11*	11.16**	.22*	4.77**	.22**
Students	90.46	34.03	1.93	137.12	4.23	67.34	3.31
Model fit statistics							
Deviance	2669.70	2354.76	1243.59	2749.79	1486.71	2510.61	1462.00
Number of parameters	2	2	2	2	2	2	2

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

Table 3
Grade 5 HLM model estimates.

Fixed effects	Gray posttest	Signaling posttest	Comparison main idea quality	Comparison total recall	Comparison recall competency rating	Problem & solution total recall	Problem & solution recall competency rating
Fixed effects							
Intercept	33.47*** (.53)	16.99*** (.28)	3.29*** (.13)	30.05*** (1.44)	4.52*** (.15)	20.31*** (.89)	3.94*** (.15)
Female	−.10 (.67)	1.47* (.57)	.24 (.13)	2.39 (1.72)	.31 (.23)	1.45 (1.34)	.19 (.20)
Gray pretest	.50*** (.06)	.18*** (.03)	.03*** (.01)	.28*** (.08)	.05** (.01)	.32*** (.06)	.08*** (.01)
Comparison recall competency pretest	.76** (.22)	–	–	2.77*** (.36)	.28*** (.07)	1.41*** (.25)	.26*** (.06)
Signaling pretest	–	.39*** (.06)	–	–	–	–	–
Comparison	–	–	.17** (.06)	–	–	–	–
Class average gray pretest	.70** (.17)	.39*** (.07)	.000 (.03)	.48 (.32)	.06 (.03)	.40 (.22)	.09* (.04)
Class average comparison recall competency pretest	−.05 (.92)	–	–	1.30 (1.89)	.03 (.19)	1.32 (1.18)	.32 (.21)
Class average signaling pretest	–	−.13 (.15)	–	–	–	–	–
Class average comparison	–	–	.53 (.28)	–	–	–	–
ITSS	2.10* (.95)	3.12** (.71)	1.14*** (.22)	9.59* (3.39)	1.19** (.31)	4.88* (1.74)	.80* (.30)
Random effects (Variances of)							
Classrooms	1.90	.05	.24***	41.51***	.20*	14.94***	.35**
Students	86.38	35.66	1.89	219.00	4.81	118.43	4.26
Model fit statistics							
Deviance	2552.18	2337.89	1266.37	2938.32	1582.68	2719.25	1554.54
Number of parameters	2	2	2	2	2	2	2

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

fidelity. The criteria for inclusion included the teacher reviews of the student performance reports, teacher observed to be interacting with students and responding to questions, and meeting the prescribed time on task (30–45 min each week for 6–7 months). We hypothesized that fidelity of implementation of ITSS would improve reading comprehension as measured by standardized test scores and researcher-designed measures.

6.1. Empirical contribution

The primary research question was whether 4th and 5th grade classrooms using the ITSS delivery of the structure strategy with fidelity as a partial substitute for the standard language arts curriculum outperform control classrooms on a standardized measure and researcher designed measures of reading comprehension?

Our analysis shows that 4th and 5th grade classrooms using ITSS with high fidelity (time used, teacher monitoring, and teacher engagement in learning environment) showed moderate to large effects on researcher-designed measures. The effect sizes on the standardized test were small, but quite impressive for a standardized reading comprehension test using a multiple choice format, particularly for the 4th graders. The effect sizes were larger than those reported when all 4th and 5th grade classrooms were used in the analysis regardless of the fidelity of implementation factors (Wijekumar et al., 2012).

6.2. Practical contribution

Improving reading comprehension is a critical goal for most schools. Pressley et al. (1998) reported that teachers in 4th and 5th grade do not provide necessary reading strategy instruction. Our research team's review of reading curricula shows little attention to text structure in

Table 4
Student and class level means, standard deviations, statistical results, and effect size of HLM analysis on adjusted scores of the Gray Silent Reading Test (GSRT; Wiederholt & Blalock, 2000).

	ITSS			Control			*t*-Test[a]
	n	*M*	*SD*	*n*	*M*	*SD*	
Grade 4							
Class level							
Pretest	12	21.96	2.94	11	22.30	2.22	$t(21) = -.31, p = .76$
Posttest	12	29.63	3.38	11	25.59	2.50	$t(17) = 3.18, p = .006$
Student lvl							
Pretest	244	21.94	10.59	223	22.30	10.71	
Posttest	224	29.58	11.97	213	25.52	10.90	
Grade 5							
Class level							
Pretest	12	26.91	3.46	9	25.08	4.75	$t(19) = 1.03, p = .32$
Posttest	12	34.93	2.88	9	31.38	4.28	$t(15) = 2.20, p = .044$
Student lvl							
Pretest	251	26.99	10.95	192	25.04	11.73	
Posttest	225	35.14	10.18	188	31.43	12.55	

[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.

Table 5
Class level means, standard deviations, statistical results, and effect sizes of HLM analyses about the comparison text.

Measure	ITSS			Control			t-Test[a]
	n	*M*	*SD*	*n*	*M*	*SD*	
4th Grade							
Signaling test	12	17.04	2.57	11	12.27	1.66	$t(17) = 5.98$, $p = .000$
Main idea quality	12	3.41	.78	11	2.23	.42	$t(17) = 6.64$, $p = .000$
Total recall	12	22.14	5.90	11	18.61	4.42	$t(17) = 1.90$, $p = .074$
Comparison competence	12	3.95	.84	11	3.21	.67	$t(17) = 2.40$, $p = .028$
5th Grade							
Signaling test	12	18.47	1.62	9	14.34	2.40	$t(15) = 4.42$, $p = .001$
Main idea quality	12	3.81	.51	9	2.69	.62	$t(15) = 5.20$, $p = .000$
Total recall	12	34.49	7.45	9	22.90	7.85	$t(15) = 2.83$, $p = .013$
Comparison competence	12	5.05	.55	9	3.63	.97	$t(15) = 3.85$, $p = .002$

[a] *t*-Test for ITSS–Control difference on posttest scores adjusted for covariates from HLM model.

reading comprehension textbooks. Interventions that provide such reading strategy instruction are urgently needed and many have been created. Unfortunately, many interventions do not scale up to produce the expected improvements in larger scale studies.

ITSS has been tested at large scale and produced some desired effects. Wijekumar et al. (2012) report on 4th grade studies that show the following effect sizes: comparison signaling test = .28, comparison main idea quality = .49, comparison total recall = .11, and comparison competency = .18. In this study of *high fidelity classrooms*, the effect sizes were: comparison signaling test = .71, comparison main idea quality = .83, comparison total recall = .26, and comparison competency = .31. Wijekumar et al., (2013) report on 5th grade studies with effect sizes of: comparison signaling test = .42, comparison main idea quality = .53, comparison total recall = .32, and comparison competency = .26. Similar to the findings at the 4th grade level, the 5th grade classrooms that implemented ITSS with fidelity reported higher effect sizes: comparison signaling test = .45, comparison main idea quality = .73, comparison total recall = .54, and comparison competency = .50.

This analysis of schools and classrooms that implemented the ITSS intervention with fidelity shows that the effect sizes can be higher. This finding suggests that the teachers' role in the implementation of such web-based learning environment cannot be discounted.

The number of classrooms that implemented the intervention with fidelity was 24 out of the 259 that participated in the large scale study. We have identified six possible reasons for this number based on our findings, classroom observations, and conversations with teachers during professional development sessions.

First, the focus of the large-scale efficacy study was on delivering the structure strategy using the web-based ITSS and therefore the teacher professional development was not extensive. We expected that the teacher's role was to monitor progress and respond to student questions without actively delivering the intervention within the classroom (e.g., teacher-led activities to write a main idea). Thus, the research team provided only one 2-h professional development session on ITSS at the beginning of the research study. We also wanted to ensure that the intervention was not dependent on the teachers' understanding of the structure strategy. At pretest when teachers were asked whether they used text structure in their classrooms. Over 70% said no and those that did use it were only aware of problem and solution or comparison text structures (only two of the five text structures covered by the structure strategy). There were 12 classrooms where the school had adopted a curriculum with some components of text structure but they were not included in this analysis. A solution to this problem is a stronger teacher professional development session followed by classroom coaching sessions to help teachers successfully use the system.

Second, teachers implementing ITSS had varying degrees of interest in new innovations. Observations of classrooms showed that 40–50% of teachers allowed students to use the software without active monitoring. In these situations teachers sat in the computer labs and graded papers or worked on other projects while students were using the software. One teacher commented that ITSS required more effort because students sometimes raised their hand and asked for help (when I.T. prompted them to do so). The teacher commented that another software application used in her school was easier to implement since students loved to "play" on the software and did not require any oversight. Again, a stronger professional development convincing the teachers about the utility of the structure strategy may sway teachers to adopt the ITSS software with fidelity. Teachers need to see the value of the software and experience the benefits of using it to be convinced of its efficacy. It is important to note that even though only 24 classrooms adopted the system with fidelity, over 100 classrooms continued to use the software beyond the research period because teachers started to observe the improvements in their students after 1-year of using ITSS. A second year of implementation may have produced a much higher number of teachers implementing the system with fidelity.

Table 6
Class level means, standard deviations, statistical results, and effect sizes of HLM analyses about the problem and solution text.

Measure	ITSS ($n = 12$)		Control ($n = 11$)		t-Test[a]
	M	*SD*	*M*	*SD*	
4th Grade					
Total recall	16.07	3.66	13.54	3.31	$t(17) = 2.59$, $p = .019$
Problem/solution competence	3.35	.83	2.77	.55	$t(17) = 2.43$, $p = .027$
5th Grade	ITSS ($n = 12$)		Control ($n = 9$)		
Total recall	23.24	5.11	15.89	5.05	$t(15) = 2.81$, $p = .014$
Problem/solution competence	4.45	.92	3.15	.91	$t(15) = 2.65$, $p = .019$

[a] *t*-Test for ITSS–Control difference on posttest scores adjusted for covariates from HLM model.

Table 7
Effect sizes of its on all reading measures 4th and 5th grades.

Measure	Coefficient for ITSS from HLM	Pooled student-level standard deviation	Effect size
4th Grade			
GSRT	3.79	10.65	.36
Comparison text			
Signaling test	4.48	6.33	.71
Main idea quality	1.26	1.52	.83
Total recall	3.62	14.13	.26
Comparison competence	.71	2.26	.31
Problem and solution text			
Total recall	3.11	10.12	.31
Problem/solution competence	.62	2.16	.29
5th Grade			
GSRT	2.10	11.29	.19
Comparison text			
Signaling test	3.12	7.01	.45
Main idea quality	1.14	1.55	.74
Total recall	9.59	17.71	.54
Comparison competence	1.19	2.41	.50
Problem and solution text			
Total recall	4.88	12.59	.39
Problem/solution competence	.80	2.40	.33

Note. Effect size = Adjusted difference between ITSS and Control groups divided by the student-level pooled standard deviation.

Third, some teachers (~25%) were very confident that their existing practices were far superior to any new approach that may be introduced. For example, one teacher was observed reminding students to use the first sentence of the passage as their guide to writing the main idea. The teacher also reminded students to re-read and look for important ideas in the text. This occurred during the middle of the academic year when students had already received instruction in writing a good main idea using the patterns presented in ITSS (e.g., the problem is, _____ and the solution is _____). Students receiving conflicting signals from the teacher and ITSS will probably defer to the teacher who is the authority figure in the classroom. Again, teachers' understanding of the value and utility of ITSS is key to overcoming these issues.

Fourth, in a small number of classrooms (<10%) there were challenges with computer lab availability and frequent outages and quality of computers. Teachers in these classrooms were overwhelmed and frustrated with handling the challenges with the computers. The research team diligently tried to support the teachers by installing software or speaking with administrators about scheduling computer lab time. Unfortunately, in many instances these issues only came to light when the research team reviewed the ITSS usage logs. Teachers rarely notified the research team on their own and as a result students sometime did not use the ITSS system for 2–4 weeks. Solving this problem continues to be a challenge as schools continue to upgrade their hardware, operating systems, and software changes causing new learning challenges for teachers and developers of software as well.

Fifth, findings from this study also suggest that implementation of educational technologies may be affected by fidelity of implementation relying on teacher reviews of student progress and teacher roles in monitoring the software use. Students who knew that the teachers were actively reviewing his/her responses and gaming history may have been more likely to work harder on the tasks and limit their gaming behaviors.

Finally, schools are under pressure to meet Federal and State mandates for student performance. The pressure to meet the prescribed achievement benchmarks make schools reluctant to participate in year-long studies that take instructional time to devote to new developments. Therefore, some large scale studies have fidelity issues due to practical time limitations in the schools. Meyer et al., (2010) conducted a study with over twice the dosage in the development studies with 90 min a week spread over two or three sessions a week and produced large effect sizes. Based on the positive results with higher dosage (Meyer et al.) and the current study it is conceivable that ITSS when implemented with high fidelity, may produce even stronger results if schools were able to devote more instructional time.

6.3. Theoretical contribution

This study contributes to the growing evidence of the effectiveness of the structure strategy in improving content area reading comprehension as measured by standardized and researcher designed measures. The structure strategy focuses on the selection and encoding of text to create tightly organized strategic memory representations that foster efficiencies in storage and retrieval. The process of selecting important ideas from the text is assisted by finding the text structure organizing the passage. Students learn to classify the organizing text structure as one of the five identified by Meyer (1975, 1985; Comparison, Problem and Solution, Cause and Effect, Sequence, and Description) using the affordances of signaling words. Students learning these text structures using the web-based ITSS showed statistically significant improvements in reading comprehension on a standardized test that used a multiple choice format with questions ranging from understanding paraphrases to drawing inferences and applications. Effect sizes were higher, as expected, for researcher-designed tasks, such as signaling words and main idea tasks for both comparison and problem and solution text structure passages.

The design of the web-based instructional strategies in ITSS with the focus on modeling, practice, assessment, and advanced feedback also provide evidence of effectiveness of technologies to deliver reading comprehension instruction, specifically structure strategy instruction. These findings for the effectiveness of technology for the delivered of complex instruction are similar to the findings by Ponce, Lopez, and Mayer (2012). Further, the evidence presented here showcases how the fidelity of implementation can contribute to higher effect sizes in reading comprehension outcomes.

In summary, the results from the current study support our theory of change (Fig. 1). The structure strategy delivered to students using effective instructional practices and presented in a web-based intelligent tutoring system can achieve stronger results in content area reading comprehension if implemented with high fidelity.

6.4. Limitations and future directions

The results of this study apply only to classrooms and schools with similar profiles as reported in the description of the participating schools. These findings are based on a random assignment of a volunteer sample of schools. We selected the whole school in which the experimental classes/teachers met our definition of high-fidelity implementation. Because classes/teachers were randomly assigned to treatment and control conditions within schools, teacher qualities within the same school should be comparable. The purpose of this study was to evaluate the effects of ITSS when the ITSS lessons were used as intended. The selection of schools with high-fidelity experimental classes/teachers was intentional and necessary to address the research question posed. Results from this study may not generalize to classes/teachers that are dissimilar to those included in this study.

Future studies may use additional approaches to measure teacher knowledge of text structure and interest in computer-based learning environments in addition to the measures of fidelity of implementation used in this research study. Some researchers have extended data collection to a second year allowing teachers' a full year to receive extended professional development and in-school coaching to understand the complexities of the intervention and be able to deliver it with confidence. These approaches may further strengthen fidelity of implementation. Further studies that replicate these findings will be necessary to draw causal conclusions.

In conclusion, this research demonstrates that the web-based intelligent tutoring system for the structure strategy can be effective in improving content area reading comprehension in grades 4 and 5. Further, teachers who implement the intervention with fidelity can expect to achieve better outcomes.

Acknowledgment

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.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, Lori Johnson, Kathryn Shurmatz, Janessa Weaver, Yu-Chu Lin, Melissa Ray, Michael Cook.

References

Armbruster, B., Anderson, T., & Ostertag, J. (1987). Does text structure/summarization instruction facilitate learning from expository text? *Reading Research Quarterly, 22*, 331–346.
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.
Bartlett, B. J. (1978). *Top-level structure as an organizational strategy for recall of classroom text*. Unpublished doctoral dissertation. Arizona State University.
Carrell, P. L. (1985). Facilitating ESL reading by teaching text structure. *TESOL Quarterly, 19*, 727–752.
Chi, M. T. H., Feltovich, P. J., & Glaser, R. (1981). Categorization and representation of physics problems by experts and novices. *Cognitive Science, 5*, 121–152.
Cook, L. K., & Mayer, R. E. (1988). Teaching readers about the structure of scientific text. *Journal of Educational Psychology, 80*, 448–456.
Drummond, K., Chinen, K., Duncan, T. G., Miller, H. R., Fryer, L., Zmach, C., et al. (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 June 26, 2012, from http://ies.ed.gov/ncee/edlabs/regions/northeast/pdf/REL_20104035.pdf.
Englert, C. S., & Hiebert, E. H. (1984). Children's developing awareness of text structures in expository materials. *Journal of Educational Psychology, 76*, 65–74.
Gartner, Inc. (2012). *Hype cycles*. Retrieved on Sept 2012 from http://www.gartner.com/technology/research/hype-cycles/.
Gordon, C. J. (1990). Contexts for expository text structure use. *Reading Research and Instruction, 29*, 55–72.
James-Burdumy, S., Mansfield, W., Deke, J., Carey, N., Lugo-Gil, J., Hershey, A., et al. (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.
Mayer, R. E. (2009). *Multimedia learning* (2nd ed.). New York, NY: 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., 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., 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., Talbot, A. P., Poon, L. W., & Johnson, M. M. (2001). Effects of structure strategy instruction on text recall of older African American Adults. In J. L. Harris, A. Kamhi, & K. Pollock (Eds.), *Literacy in African Americans communities* (pp. 233–263). Mahwah, NJ: Erlbaum.
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). (2011). Retrieved July 1, 2011 from http://nces.ed.gov/nationsreportcard/pdf/main2011/2012459.pdf.
Paris, S. G., Cross, D. R., & Lipson, M. Y. (1984). Informed strategies for learning: a program to improve children's reading awareness and comprehension. *Journal of Educational Psychology, 76*, 1239–1252.
Polley, R. R. (1994). *Facilitating recall through awareness of text structure*. Unpublished masters thesis. Lexington: University of Kentucky.
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*(2012), 1170–1183.
Pressley, M., Wharton-McDonald, R., Mistretta-Hampton, 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.

Raphael, T. E., & Kirschner, B. W. (1985). *The effects of instruction in comparison/contrast text structure on sixth-grade students' reading comprehension and writing products* (Research Series No. 161). East Lansing: Michigan State University, Institute for Research on Teaching.
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.
Richgels, D. J., McGee, L. M., Lomax, R. G., & Sheard, C. (1987). Awareness of four text structures: effects in recall of expository text. *Reading Research Quarterly, 22*, 177–196.
Samuels, S. J., Tennyson, R., Sax, L., Mulcahy, P., Schermer, N., & Hajovy, H. (1988). Adults' use of text structure in the recall of a scientific journal article. *Journal of Educational Research, 81*, 171–174.
Weisberg, R., & Balajthy, E. (1989). Transfer effects of instructing poor readers to recognize expository text structure. In S. McCormick, & J. Zutell (Eds.), *Cognitive and social perspectives for literacy research and instruction: Thirty-eighth yearbook of the national reading conference* (pp. 279–285). Chicago: National Reading Conference.
Wiederholt, J. L., & Blalock, G. (2000). *Gray silent reading tests (GSRT)*. Austin, TX: Pro-Ed.
Wijekumar, K., Hitchcock, J., Turner, H. W., Lei, P., & Peck, K. (2009). *A multisite cluster randomized trial of the effects of CompassLearning Odyssey® Math on the math achievement of selected grade 4 students in the Mid-Atlantic Region*. NCEE 2009-4068 http://www.eric.ed.gov/PDFS/ED507314.pdf.
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.
Wijekumar, K., Meyer, B. J. F., Lei, P., Lin, Y., Johnson, L. A., Shurmatz, K., et al. (2013). *Improving reading comprehension for 5th grade readers in rural and suburban schools using web-based intelligent tutoring systems*. Paper submitted for publication.
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.