Journal of Educational Psychology
2011, Vol. 103, No. 1, 140–168

© 2011 American Psychological Association
0022-0663/11/$12.00 DOI: 10.1037/a0021606

Individualizing a Web-Based Structure Strategy Intervention for Fifth Graders' Comprehension of Nonfiction

Bonnie J. F. Meyer
The Pennsylvania State University at University Park

Kausalai K. Wijekumar
The Pennsylvania State University at Beaver

Yu-Chu Lin
The Pennsylvania State University at University Park

In the study, we investigated effects of 2 different versions of a web-based tutoring system to provide 5th-grade students with strategy instruction about text structure, which was an intervention to improve reading comprehension. The design feature assessed varied in individualization of instruction (individualized or standard). The more individually tailored version was developed to provide remediation or enrichment lessons matched to the individual needs of each student. Stratified random assignment was used to compare the effects of 2 versions of the 6-month web-based intervention. Students in the individualized condition made greater improvements from pretest to posttest on a standardized reading comprehension test ($d = 0.55$) than did students in the standard condition ($d = 0.30$). Students receiving more individualized instruction demonstrated higher mastery achievement goals when working in the lessons than did students receiving the standard instruction ($d = 0.53$). Students receiving more individualized instruction showed greater improvement in using signaling, better work in lessons, and more positive posttest attitudes toward computers than did students receiving standard instruction. Students in both conditions improved their recall of ideas from texts and their use of the text structure strategy and comparison signaling words.

Keywords: reading comprehension instruction, web-based tutoring, individualization, structure strategy, fifth-grade reading

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

Adapting learning situations to match students' aptitudes is a critical issue for learning theorists (e.g., Vygotsky, 1978), educational psychologists (e.g., Corno et al., 2002), instructional designers (e.g., Camp, Paas, Rikers, & van Merrienboer, 2001; Kalyuga & Sweller, 2005), developers of computer-based tutors (e.g., Graesser et al., 2004; Woolf, 2009), and teachers (e.g., Corno, 2008; Gregory & Chapman, 2002). Historically, individualization has been a particular interest of literacy instruction and research (e.g., Blok, Oostdam, Otter, & Overmaat, 2002; Chall & Dale, 1995; Dale & Chall, 1948; Gambrell, Morrow, & Pressley, 2007; Reutzel, 2007).

Attempts to provide individualized reading activities within a classroom can take a toll on teacher effectiveness due to management demands (Reutzel, 2007). A major strength of web-based reading comprehension instruction is the delivery of individualized timely feedback to support student learning (e.g., Dalton & Strangman, 2006; McNamara, O'Reilly, Rowe, Boonthum, & Levistein, 2007; Meyer & Wijekumar, 2007; Meyer et al., 2010). The learner-adapted version of a web-based tutoring program, investigated in this study, may be particularly helpful to classroom teachers because it could potentially improve their ability to differentiate literacy instruction without overwhelming them with extra management loads, which can be off-loaded to the tutoring program. The learner-adapted version enables individualization and provides the opportunity for both remediation and enrichment.

In this study, we compared reading comprehension instruction adapted to fifth-grade students' online performance with a standard sequence of instruction. The instruction focused on using the structure strategy with nonfiction texts (e.g., expository and persuasive texts). The structure strategy teaches students to use text structures to increase comprehension. For example, students learn about signaling words (*in contrast*, *solution*), which can clue readers into arguments often made in expository texts (e.g., Meyer, Brandt, & Bluth, 1980). Meyer et al. (2002) established the ben-

Bonnie J. F. Meyer and Yu-Chu Lin, Department of Educational and School Psychology and Special Education, The Pennsylvania State University at University Park; Kausalai K. Wijekumar, Information Science and Technology, The Pennsylvania State University at Beaver.

The opinions expressed are those of the authors and do not represent views of the Institute or the United States Department of Education. More information on our project is available at http://itss.br.psu.edu/. The research reported here was supported by Institute of Education Sciences, United States Department of Education Grant R305G030072 to The Pennsylvania State University at University Park. We appreciate the contributions of the elementary students, faculty, and administration involved in this research effort. Additionally, we appreciate the input of other faculty, affiliates, and students at The Pennsylvania State University, including James Spielvogel, Rayne Sperling, Pui-Wa Lei, Catherine Meier, Kelli Higley, Cynthia Spencer, Barbara Shochowitz, Melissa Ray, Kathryn Shurmatz, and Lori Johnson.

Correspondence concerning this article should be addressed to Bonnie J. F. Meyer, Department of Educational and School Psychology and Special Education, 204 Cedar Building, The Pennsylvania State University at University Park, University Park, PA 16802. E-mail: bjm8@psu.edu

efits of the web-based structure strategy intervention with fifth-graders over a randomized control condition. In the current investigation, we sought to determine whether reading comprehension and intrinsic motivation in the lessons increased when the intervention matched the type of practice lessons and the difficulty of the text within lessons to the individual needs of each student.

Theoretical and Historical Background of the Structure Strategy Intervention

Research about the structure strategy has grown out of cognitive research on reading comprehension that focused on memory representations and recall after reading (e.g., Crothers, 1972; Kintsch & van Dijk, 1978; Mandler & Johnson, 1977; Meyer, 1975; Meyer, Brandt, & Bluth, 1980; Meyer & McConkie, 1973). In addition, structure strategy instruction reflects current comprehension models (e.g., van den Broek, Rapp, & Kendeou, 2005). Van den Broek's comprehension model (van den Broek, Young, Tzeng, & Linderholm, 1999) points to the importance of top-down processes aimed at seeking coherence (strategic, goal-directed searches for meaning, e.g., Graesser, Singer, & Trabasso, 1994). The structure strategy is one such strategy that is focused on seeking coherence among text ideas.

Readers who use the structure strategy (Meyer, 1985; Meyer et al., 1980; Meyer & Poon, 2001) approach reading with the strategic knowledge that authors structure texts in predictable ways and that they can construct an integrated representation of a text by following the hierarchical organization of the text and the relative importance of its conceptual content. When reading a text, a reader builds this mental representation of the information in the text (Britton & Graesser, 1996; Gernsbacher, Varner, & Faust, 1990; Kintsch, 1998). One important way to build the coherent mental representation needed for encoding and retrieving information from a text is to use the structure of the text (Grimes, 1975; Mann & Thompson, 1986; Meyer, 1975; Sanders, Spooren, & Noordman, 1992). Due to limited processing capacity (Baddeley, 1992), readers cannot remember and learn everything in a text, so some information must be selected for deeper encoding over other information. Focusing on the superordinate or top levels of the structure of the text (top-level structure [TLS]) can help readers select the most important information for thorough encoding. This approach helps readers understand the logical structure used by an author to communicate important ideas. Readers who use the structure strategy tend to understand and retain more of what they read than those who do not use the strategy (e.g., Meyer, Young, & Bartlett, 1989). Instruction about text structure has yielded positive effects on comprehending and remembering information from texts with children, young adults, and older adults (e.g., Armbruster, Anderson, Ostertag, 1987; Bartlett, 1978; Cook & Mayer, 1988; Meyer et al., 2002; Meyer & Poon, 2001; Meyer et al., 1989; Raphael & Kirschner, 1985; Williams, Stafford, Lauer, Hall, & Pollini, 2009).

The materials and instruction used in the current study were based on analyses of the reading performance of hundreds of good readers across the life span interacting with expository texts (Meyer et al., 1980; Meyer & Rice, 1984, 1989). In addition, instruction followed best practices in literacy instruction, such as the gradual release of responsibility model of instruction (see Reutzel, 2007), and included explicit explanation, modeling, and scaffolding of the strategy.

The current investigation extended previous research in the use of web-based structure strategy instruction. Meyer et al. (2002) developed the first web-based delivery of the structure strategy. The structure strategy instruction used by Meyer et al. (2002) was adapted from lessons used in earlier randomized control studies (Meyer & Poon, 2001; Meyer, Young, & Bartlett, 1989) that demonstrated that learning the structure strategy substantially increased reading comprehension over two types of control groups per study: a no-contact control and a control group that read the same texts as the structure strategy group, but with a practice strategy or a motivational strategy.

In the Meyer et al. (2002) study, fifth-grade students were randomly assigned to two web-based versions of the structure strategy or a control (regular classroom reading activities). This study showed significant improvements in recall and structure strategy use by fifth-grade students after use of a web-based interface for presenting information and an e-mail based interface for interacting with older adult tutors. Superiority in the amount of information remembered from text by the structure strategy group over the control group in this randomized control group design was found 2 1/2 months after training. The average reader receiving tutoring with the structure strategy training had a total recall score equal to a reader in the control group who scored at the 81st percentile on the delayed posttest ($d = 0.92$). Additionally, general self-efficacy improved from pretest to posttest for the group with human tutors, in comparison with the control group and a group with the web-based instruction without tutors. No changes were found in attitudes toward computers among groups. Improvement was needed in the delivery of the system because it was cumbersome for tutors and students to use and did not allow for immediate feedback.

Human tutors can be costly and inconsistent and can limit widespread accessibility. As a result, Meyer and Wijekumar (2007) further adapted and improved the earlier web-based approach. Improvements included the (a) introduction of an animated tutor named Intelligent Tutor (I.T.) who could provide immediate feedback, (b) addition of audio and more lessons, and (c) design of a more sophisticated delivery system, called Intelligent Tutoring of the Structure Strategy (ITSS). The agent's feedback in ITSS was modeled after the web-based interactions of fifth-graders and adult tutors who worked asynchronously online with the same structure strategy lessons (Meyer et al., 2002).

Meyer et al. (2010) examined the effects of two design features varied for ITSS on fifth- and seventh-grade students' reading comprehension. The two features were the type of feedback provided by the web-based tutor (elaborated vs. simple feedback) and the motivational factor of choice of text topics in practice lessons (student choice of texts versus no choice). In the elaborated feedback version, the animated agent provided advanced, elaborated feedback with scaffolding to improve performance on subsequent trials, whereas the other feedback version involved the animated agent providing only simple feedback about the correctness of a student's response. For example, in response to the same student's performance, the tutor in the elaborated feedback version said, "Your structure, main idea, and details are correct. Great job! But your signaling words were incorrect. Using the chart as your guide, rewrite the signaling words," whereas the tutor in the simple feedback version said, "Try again." On a third deficient trial,

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

students in the elaborated feedback group were given a model response to correct their performance writing main ideas. This type of elaborated feedback was not provided for students in the simple feedback condition.

Students who received elaborated feedback performed better on a standardized test of reading comprehension than did students who received simple feedback. Choice between two topics for practice lessons did not increase reading comprehension on the standardized test. Substantial effects sizes were found from pretest to posttest on various measures of reading comprehension. Maintenance of performance over summer break (4 months) was found for most measures.

Does Increased Individualization Increase Reading Comprehension?

The development of structure strategy instruction with an animated tutor, I.T. (see Figure 1) provided immediate feedback and increased the potential accessibility of structure strategy instruction. The elaborated feedback condition of the Meyer et al. (2010) study was individualized in that students progressed through the lessons at their own pace and the animated agent's feedback and scaffolding depended on the responses of the student. However, the order of the lessons, topics, and readability levels of the texts within the lessons were the same for all students. In the current study, we sought to further improve ITSS by individualizing the instruction (lesson content) to better meet the needs of fifth-grade students with diverse skills in reading.

Comprehension support for online reading of expository texts frequently has included easier versions of texts (i.e., Leong, 1995; Reinking, 1988; Reinking & Schreiner, 1985). These easier versions have substituted more frequently used words for low-frequency words, less technical words for more technical words, or more concrete words for more figurative words. The effectiveness of such easier versions for online reading of fifth-grade students has been inconclusive. Leong (1995) reported no facilitation of the easier versions for reading comprehension, whereas Reinking (1988) found increased reading comprehension from easier versions. Use of easier versions of texts was one of several ways that we adapted individualized ITSS to meet the instructional needs of learners. (The last column in Appendix A shows the availability of easier texts for lesson content to meet the individual needs of students during instruction with individualized ITSS. Appendix A also lists the lessons in standard ITSS in their fixed order as well as alternative text topics available for individualizing instruction in individualized ITSS.)

The individualized version of ITSS attempted to accomplish two goals, providing assistance for students who were having difficulty (moderate and more severe) and supplying enrichment for students

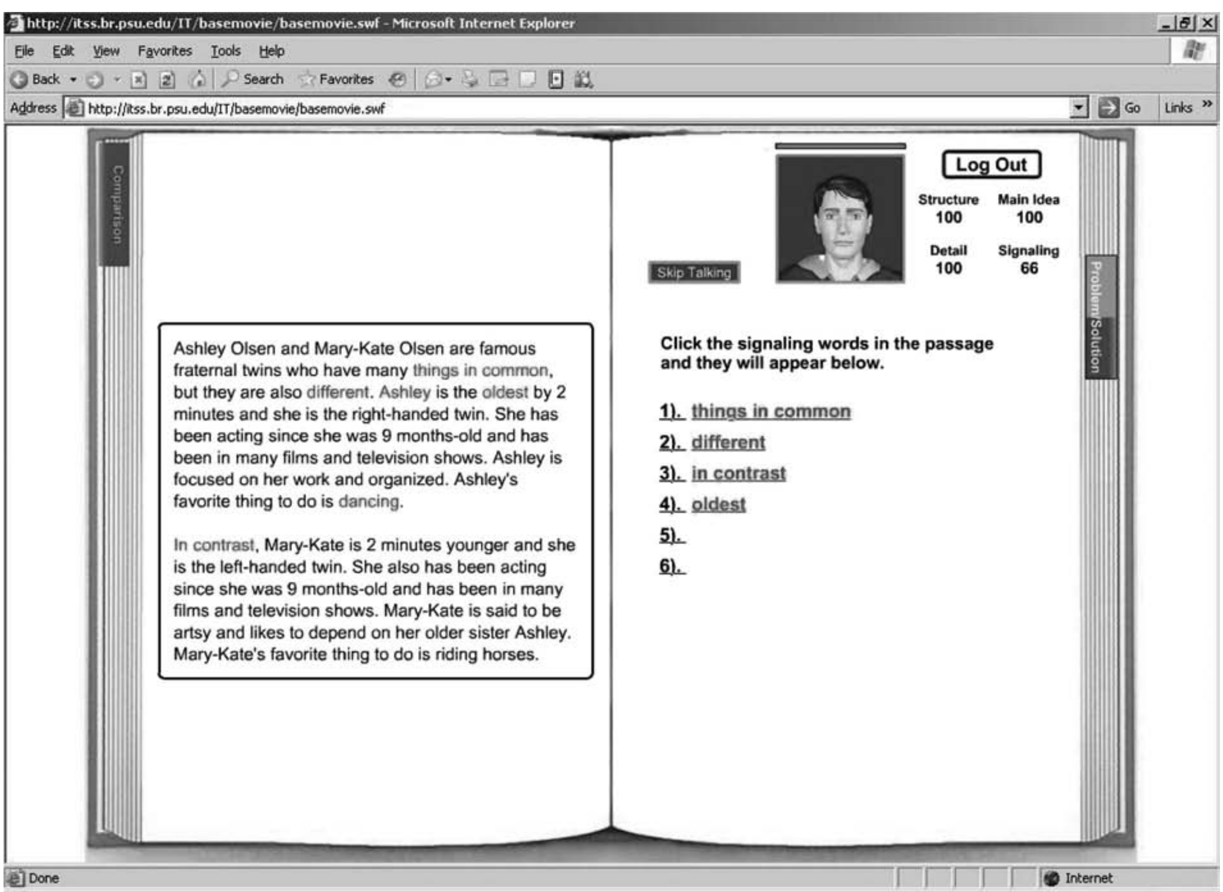

Figure 1. Sample screen for review lesson of comparison signaling (Lesson 23 in Appendix A).

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

who were demonstrating good progress in using the strategy. Based on a student's performance in the current lesson, the learner-adapted instruction placed the student into an appropriate next lesson rather than following the standard sequence of lessons. As shown in the first two columns of Appendix A, standard ITSS has 65 lessons. As seen in Table 1 (see Practice row), 29 of these lessons focus on practicing the strategy. Appendix A shows the alternate text topics for practice lessons, which were used in individualized ITSS in order to adapt instruction to the reading needs of the student. Additionally, we identified 40 texts from ITSS practice, review, and integration lessons (see Table 1) that had readability levels at fifth-grade or higher. For each of these texts, we prepared easier versions for use in individualized ITSS (see Appendix A).

In the more learner-adapted version of ITSS, a student's next lesson may provide *remediation* by giving the student more practice with the same objectives as the previous lesson but with a new text containing (a) similar length, structure, and readability or (b) similar structure and easier readability. This procedure gives students more practice with a particular lesson's objectives without the stigma of appearing to repeat the lesson. It also provides more cognitive support through the repetition of objectives, structure, and signaling words, which was expected to increase familiarity with these lesson elements and learning. Alternatively, the student's next individualized lesson may provide *enrichment* by delivering new lesson content with less familiar topics and more difficult text than the equivalent lesson in the standard version.

The approach to learner-adapted instruction was relatively simple and as a result quickly provided adapted instruction to many students. We applied a dynamic online assessment of students' use of the structure strategy by measuring their performance in recalling the main ideas from a text during free recall, the final task in a lesson. Use of the structure strategy has been shown to increase the number of main ideas produced in a free recall (Meyer et al., 1989). Rapid diagnostic tests for real-time monitoring of reading and strategy performance of each child do not require the complex continuous tracing of student models used in many intelligent tutoring systems (e.g., Graesser et al., 2004). Various approaches of learner-adapted instruction in e-learning environments have pointed to their effectiveness over nonadapted instruction (e.g., Camp et al., 2001; Kalyuga & Sweller, 2005), but most have tested well-structured domains, such as mathematics. Our rapid online assessment was similar to that used by Kalyuga and Sweller (2005) with tenth-grade students using an algebra tutor (performance on an equation, such as $-3x = 7$) but had less constrained problems and answers. Similar to the recommendations by other researchers (e.g., Aleven, Ashley, Lynch, & Pinkwart, 2008), Kalyuga and Sweller's (2005) article called for doing research in less structured areas and specifically monitoring language comprehension in online reading tutors. Our research addresses this need.

As seen in Table 2, individualized ITSS did not provide students with more time in ITSS, lessons, or texts to read than did standard ITSS. Instead, individualized ITSS better matched the practice lessons a student received to their immediate needs based on their performance in the current lesson. When individualized ITSS detected difficulties in a student's understanding during a practice lesson, in the subsequent lesson, jumps in complexity were reduced. In standard ITSS, no adaptations occurred from the stan-

Table 1
Order, Number, and Type of Standard ITSS Lessons by Structure

	Highlighted top-level structure (in context of other structures[e])				
Lessons	Comparison (C[f])	Problem and solution (P & S[f])	Cause and effect (C & E[f])	Sequence (S[f])	Description (D[f])
Order of lessons	1	2	3	4	5
			Total number		
Standard condition	12	12	16	12	13
(Individualized)	(19)	(16)	(20)	(18)	(20)
			Type and number of lessons in standard condition		
I.T. models strategy[a,b]	2	2	1	1	1
Practice[a]	7	4	4	7	7
Let's check	1	1	3	1	1
Review structures		1	1	1	1
Review via writing		1	1	1	1
TLS integration[c]		3	6	1	2
Taught in context of other structures[e]	d[f], c & e	C[f], C & E, d	P & S, C, d	P & S, C & E	S, C & E, C
Other[d]	2				

Note. Adapted with permission from "Web-Based Tutoring of the Structure Strategy With or Without Elaborated Feedback or Choice for Fifth- and Seventh-Grade Readers," by B. J. F. Meyer, K. Wijekumar, W. Middlemiss, K. Higley, P. Lei, C. Meier, and J. Spielvogel, 2010, *Reading Research Quarterly, 45,* p. 64. Copyright 2010 by the International Reading Association. ITSS = Intelligent Tutoring of the Structure Strategy; I.T. = Intelligent Tutor; TLS = top-level structure.
[a] Tasks: These were signaling, structure, main idea, and recall tasks. [b] Standard and individualized groups received the same initial lessons for each structure. [c] TLS integration: These lessons use diagrams to show a text's logical structure that integrates other important structures embedded within the highlighted top-level structure. [d] Other: These early lessons involve writing titles and correcting work of other students. [e,f] Initials represent structure names: If capitalized, the structure is explicitly integrated as a substructure used in a text with the highlighted top-level structure. Lower case indicates implicit teaching of the other structure at this point in the lessons.

Table 2
Means (Standard Deviations) for Numbers of Lessons, Texts Read, and Time for Standard and Individualized ITSS

	Individualization condition					
	Standard		Individualized		Statistics	
ITSS use factors	*M*	*SD*	*M*	*SD*	*t*(129)	*p*
Number of texts read	51.52	27.23	51.72	23.25	.05	.963
Number of lessons worked	37.11	15.65	38.40	14.01	.50	.619
Number of 30-min ITSS sessions	34.53	10.44	35.06	9.51	.31	.759

Note. Standard $n = 66$. Individualized $n = 65$. ITSS = Intelligent Tutoring of the Structure Strategy.

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

dard sequence of lessons, shown in Appendix A (left column). As a result, complexity jumps were greater in terms of objectives and text characteristics. For students experiencing comprehension problems in a lesson, reduction in complexity jumps was expected to provide greater (a) learning about signaling, (b) depth of text understanding, and (c) use of the structure strategy with difficult, multicomponent text structures (i.e., problem with an embedded cause and a solution to eliminate the cause). Reduction in complexity jumps also was expected to provide an environment more conducive for developing learning achievement goals, fostering greater interest in the structure strategy lessons.

Motivation for academic achievement has often been studied with a goal orientation framework (e.g., Ames, 1992; Dweck, 1986; Meece, Blumenfeld, & Hoyle, 1988; Murphy & Alexander, 2000; Pintrich, 2000). Two goal orientations are typically studied, learning–mastery goals and performance–ego goals (Ames & Archer, 1988; Wentzel, 1999). Learning goals focus on developing competence, whereas performance goals focus on doing better than others (performance-approach goals) or at least not looking worse than others (performance-avoidant goals; Elliot & Church, 1997; Harackiewicz, Barron, Pintrich, Elliot, & Thrash, 2002). Mastery-oriented students enjoy the learning task and define competence through setting intrapersonal standards. Elliot and Church (1997) found that learning goals increased intrinsic motivation and performance approach goals increased graded performance.

O'Donnell, Reeve, and Smith (2007) suggested that learning goals are more likely to develop from instruction matched to a student's skill level or enabled through scaffolding rather than instruction that is too hard or easy. Without such matching and scaffolding, students can become frustrated with too easy or too difficult instructional tasks and stop trying to understand the strategy and improve their understanding. Scaffolding is central to Vygotsky's (1978) theory and can be a critical asset of web-based tutoring by offering tips, pointers, hints, and assistance, and selecting task difficulty. Such hints and assistance are provided in both individualized and standard ITSS, but selection of task difficulty is the strength of the more adaptive version of ITSS. Modifying and simplifying the standard lessons to adjust to the student's current level of performance were the focus of the individually tailored ITSS. Thus, the more individualized ITSS intervention was expected to increase learning goals but not performance goals.

Research Questions

In this study, we investigated the effects on reading comprehension of two different versions of a web-based tutoring system to teach the structure strategy to fifth-grade students. The fifth-grade level was targeted for instruction about text structure because most fifth graders are transitioning from reading mainly narrative texts to reading expository texts at school and in everyday life. In addition, previous research (Meyer et al., 2002) has established the benefits of web-based structure strategy intervention over a control.

The design feature investigated was individualization of instruction (individualized ITSS vs. standard ITSS). The primary research question was whether the design variation of individualizing lesson sequence, difficulty, and practice affected reading comprehension in comparison with the standard sequence of lessons. Measures of reading comprehension included researcher-designed assessments, as well as transfer to a standardized reading comprehension test (*Gray Silent Reading Test* [GSRT]; Wiederholt & Blalock, 2000). The current study answered the following primary research question. Did students in the more individualized ITSS perform better than students in standard ITSS on (a) comprehension measures practiced in ITSS with new topics (i.e., free recall), (b) application of comparison signaling on an unpracticed task, and (c) far transfer to a standardized test of reading comprehension?

The study addressed several secondary research questions. Did the variation in individualization of instruction affect learning goals regarding ITSS, quality of work in ITSS lessons, or attitudes toward computers or self? Were pretest to posttest gains found for remembering information, understanding signaling, and using the structure strategy after instruction with ITSS, and were these gains maintained 1 month after instruction? Did our attempt to better match online performance to the lesson sequence, text difficulty, and amount of practice similarly impact students who were initially high, middle, or low on a standardized reading comprehension test?

Students who completed the more individualized version were expected to show greater (a) learning of the structure strategy, (b) transfer to a standardized reading comprehension test, and (c) motivation to master the strategy than students who completed the standard ITSS. Higher scores for reading comprehension and intrinsic motivation (mastery achievement goals) were expected to result from the better match between learner needs and text–lesson difficulty. In other words, the intervention was predicted to have both a cognitive effect and a motivational effect.

Method

Participants

Participants were 131 fifth-grade students attending two elementary schools in a western Pennsylvania suburban school dis-

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

trict. Across the two schools, 80.6% of students were Caucasian, 11.4% were African American, 1.6% were Asian American, and 6.4% were Native American, Hispanic, or students from other backgrounds; 9.8% of all students received state aid in the form of free or reduced-rate lunch, and 8.5% of the students were enrolled in part-time special education services. The two elementary schools had similar state reading assessment scores (first elementary school: 54th percentile [*M* scale score for reading = 1,371 & scale score *SE* = 70], students with advanced reading skills = 50.7% and students with basic reading skills = 7%; the other district elementary school: 55th percentile [*M* scale score for reading = 1,388 & scale score *SE* = 71], 50% advanced and 10% basic). There were no significant differences between the two schools on measures of reading comprehension used in this study or significant interactions with treatments or time of testing, so the data were collapsed across schools.

Reading ability subgroups. Three reading comprehension subgroups resulted from dividing the participants into thirds on the basis of standardized pretest scores on the reading comprehension test (per form). The mean GSRT raw score for the high reading comprehension group was 45.77 (*SD* = 9.09), and the mean for the middle group was 34.95 (*SD* = 9.45). The mean was 23.21 (*SD* = 9.20) for the low reading ability group. According to GSRT norms, these scores corresponded to reading levels of 10th grade, sixth grade, and third grade for the high, middle, and low groups, respectively.

School context. ITSS was primarily self-standing instruction with little teacher input. The district and school administrators were supportive of ITSS and the time requirements of the program. The reading curriculum used by all of the teachers was the Harcourt Trophies reading series. The teachers did not use any instruction about text structure as part of the regular language arts or social studies curricula. The teachers and students did use technologies to summarize ideas in PowerPoint presentations as part of their regular classroom work. Classroom teachers voiced positive statements about the effects of ITSS on the students' writing in the classroom.

The teachers and students were comfortable with technology use in the classroom. When it was time for ITSS, the students went to the mobile docking station brought into their classrooms. Then, students took Dell laptop computers to their desks. Each student had an envelope that contained password information and headphones. Students then worked at their desks on ITSS until an announcement was made that the 30-min session had ended. The students would then return the computers to the mobile docking station. On the next session, ITSS started students at the lesson and page on which they previously stopped.

Structure Strategy Instruction

Overview of ITSS. The organization of ITSS lessons is provided in Table 1. Students in both conditions learned the five text structures in the order specified in the first row of Table 1. The second row in Table 1 displays the total number of lessons available per text structure for students in the standard ITSS condition. For example, all students in the standard condition who completed the comparison and problem-and-solution lessons completed a total of 24 lessons, 12 per structure. The third row in Table 1 shows in parentheses the total number of lessons available per structure in the individualized condition. The number varied per student and was dependent on an individual's performance in the lessons.

Students in both conditions received the same versions of the lessons when a new text structure was introduced and modeled by I.T. (see Table 1, row 4; one or two per structure). The fifth row in Table 1 lists the number of practice lessons for each structure for students in the standard condition. Potentially twice as many practice lessons were available to students in the individualized condition than those in the standard condition. For example, the seven standard practice lessons for the comparison structure had one alternate version for each lesson (see Appendix A). Thus, there were 14 practice lessons possible for individualized students demonstrating a need for remediation consistently throughout instruction with the comparison structure. In addition to extra practice lessons, students in the individualized condition could receive easier versions (third-grade to fourth-grade levels) of texts with readability levels at or above the fifth-grade level (e.g., Flesch, 1948). There were 40 easier versions of texts available for students in the individualized condition but no easier versions for students in the standard condition. The average Flesch-Kincaid grade level readability score[1] (Test Your Document's Readability, 2007) for the easier versions was 4.11 (*SD* = 0.85), whereas the average readability score for the original 40 texts read in the standard condition was 7.13 (*SD* = 1.89), $t(54.31) = 9.22$, $p < .0005$.

Commonalities between individualized and standard ITSS. The web-based ITSS system has a talking animated agent, interactive flash activities, parsers, spell checking, synonym checks, Penn Treebank (Penn Treebank, 2005), Wordnet (Miller, 2005), and elaborated feedback to teach students how to use the structure strategy (Meyer & Wijekumar, 2007; Meyer et al., 2010). ITSS was designed to include modeling of the structure strategy by an animated pedagogical agent (I.T.) who looks and speaks like a high school boy (see Figure 1). Voice recordings from an adolescent vocalist with a warm, encouraging manner were used for the agent's speech rather than computer generated speech. I.T. narrated essential parts of the instructional materials shown visually on the screen. A definition of structure or a new signaling word turned a contrasting color when I.T. spoke about it. For most lessons, I.T. read the articles aloud as the students read along. Then, students reread the articles by themselves before the recall task. Students could click on the first word of any article if they wanted I.T. to read the text again with them. The audio component of ITSS was designed to provide needed support for poor readers by helping them to understand the instruction. In contrast to the current study, the randomized control study (Meyer et al., 2002), which documented the benefits of the online structure strategy intervention, did not have audio.

ITSS presented students with modeling of the strategy, guided and independent practice, and feedback. ITSS provided student

[1] Flesch-Kincaid readability calculations (Test Your Document's Readability, 2007) throughout the article were based on the formula (.39 × ASL) + (11.8 × ASW) – 15.59, where ASL is the average sentence length (number of words divided by the number of sentences) and ASW is the average number of syllables per word (number of syllables divided by the number of words). The Flesch-Kincaid readability calculations in Figure 3 are noted by F-K = __.

interaction with the web-based system in the form of learning activities (e.g., click on signaling words, write a main idea, and write a recall of a passage—see Appendix A), assessment of student responses, and immediate, elaborated feedback based on the assessment. ITSS taught readers to (a) identify the overall TLS of expository texts in the context of other embedded text structures supporting the TLS (see Table 1 row 10), (b) write the main idea using specific patterns for each TLS (e.g., see Figure 2a), and (c) organize understanding and recall by using the TLS and recall pattern (see Figure 2b).

Lesson content. A goal for ITSS instruction is to increase students' understanding and memory of the content of the texts. For example, in ITSS there are eight articles of varying lengths and complexity related to the Pony Express (at least one using each of the five major TLSs). These articles include an advertisement calling for riders, an article contrasting Wild Bill Hickok and Buffalo Bill Cody, and an article about the effects of the trans-continental telegraph on the Pony Express. ITSS teaches students how to integrate content across different passages with the five basic structures and how to relate an author's purpose to the text structure(s) the author uses. The standard lessons include 134 texts ranging from 13 to 814 words. Instruction and practice lessons focus on one text. Review lessons include numerous texts per lesson for reviewing text structures and their signaling words. The average Lexile grade equivalent for the standard ITSS texts is 5.58 (SD = 2.17; range from Grade 1.5 to Grade 12; Lexile, 2005). Most texts came from authentic sources (e.g., 814-word text from a magazine for youth). Text topics (see Appendix A) include science (34%), social studies (28%), animals (23%), sports (9%) and food (6%). Variety in content, style, domain, and difficulty was designed to promote learning and transfer of the strategy.

Text structures. The five structures were presented in ITSS in the following order: comparison, problem-and-solution, cause-and-effect, sequence, and description (see Table 1). The less familiar and more difficult structures were explicitly taught first before the easier text structures of sequence and description (Englert & Hiebert, 1984; Meyer et al., 1989, 2010). Many lessons addressing a particular structure also included information pertaining to other subordinate structures found in the text. As students progressed past the first 12 lessons, the embedded nature of the other structures was made more explicit, and integration of structures was taught (see Table 1, particularly notes c, e, & f).

Organization of instruction. When a new text structure was introduced, I.T. modeled the strategy, the main idea and recall patterns, and a few signaling words. Next, students completed a practice lesson, which included identification of signaling words. The first lessons per structure provided scaffolding to help students correctly use the strategy. As a student progressed through the practice lessons for each structure, less and less scaffolding was provided.

Most of the first comparison lessons involved the identification of comparison signaling words, naming the comparison structure, using the comparison structure to write a thorough main idea, and writing information remembered from the article using the comparison structure. I.T. provided much assistance in the early lessons, such as reading the text for students and providing visual prompts for use of the strategy. Near the end of the comparison lessons, the student worked without help from I.T. except for feedback. The next lessons involved teaching the structure strategy with the problem and solution structure (see Table 1). The last lessons about the problem and solution structure involved review and integration with the comparison structure. Key components for

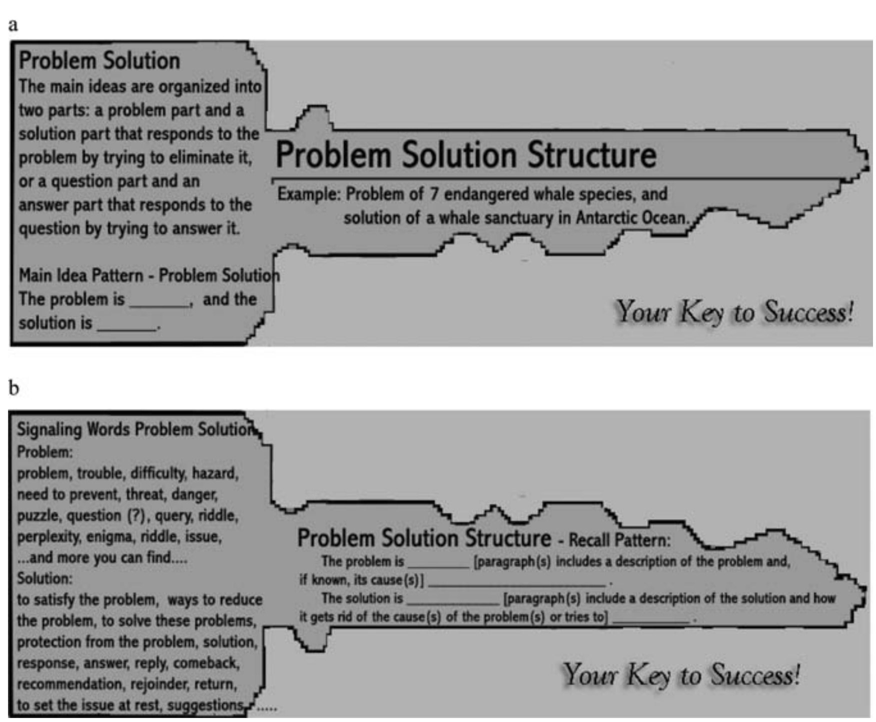

Figure 2. Problem and solution key. Front (a) displaying definition, main idea pattern, and whale example and back (b) showing signaling and pattern for writing problem and solution recall.

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

all structures involved structure specific patterns and signaling words (see Figure 2a and 2b for the problem-and-solution structure). Next in ITSS were the cause and effect lessons. These were followed by lessons focusing on the easier structures of sequence and description (see Table 1). Prior structures were reviewed and explicitly integrated with subsequent structures.

Practice. Practice tasks were designed to help students learn and apply the structure strategy (see capitalized letters [i.e., *SI*, *M*, *R*] representing types of practices tasks [i.e., *SI* = click on signaling words, *M* = write main idea, and *R* = write recall] in Appendix A). These tasks included (a) clicking on signaling words (see Figure 1): This was aided by a structure strategy key chain, which contained a laminated key for each structure with its list of signaling words (e.g., Figure 2b); students also could access an expandable signaling word table within ITSS; (b) writing the name of the structure: Beginning lessons for each structure focused on using the signaling words in the passage to identify the structure; later lessons with more complex texts required identification of the overall TLS that integrated supporting text structures to convey an author's message; (c) writing main ideas (see Figure 2a) for the text passages they read: In the first half of ITSS lessons for each structure, the student wrote the main idea while the passage was still on the screen; as students progressed through lessons about a structure, they wrote the main idea without the passage available; (d) writing a recall (see Figure 2b) of the passage: Usually students constructed a main idea earlier in the lesson and used this to aid the writing of the recall; (e) filling in a tree diagram showing the ideas being compared, the problem and its solution, the cause and its effect, or the sequence of steps in a process or time line; (f) clicking on answers for a few multiple-choice questions or answering a few open-end questions; (g) creating titles for passages based on text structure; (h) creating texts given signaling words and some general themes; and (i) correcting main ideas or recalls written by other students (provided by I.T.).

Automated ITSS scoring. The automated ITSS scoring used a parser, a dictionary for spelling, and a synonym table (populated by using data collected in all previous studies, e.g., Meyer et al., 2002, 2010). ITSS scoring focused on main ideas, details, and signaling and was based on Meyer's (1975, 1985; Meyer et al., 1989) content structure approach. The content structure is a hierarchy of text ideas organized primarily by text structures. The approach is based on Grimes's (1975) semantic grammar of propositions rather than repetition of words. The TLSs for a text provide the basis for the hierarchy and the determination of the main ideas and the details. For example, the main ideas in the problem-and-solution text shown in Appendix B are the problem (psychologists become allergic to rats), its cause (protein in the rats' urine), and the solution (kindness to rats) that helps to eliminate the cause (reduces spattering of urine on psychologists). The details elaborate on these ideas (recommendation presented at a National Institutes of Health sponsored meeting) or are tangentially related to them (Andrew Slovak is British). Articles and prepositions do not appear in the content structure; instead, the content structure is composed of words from the text representing ideas and labels showing how the ideas are related. An adapted content structure (devoid of labels and reduction of redundant words) was programmed into ITSS for each text to be read by students in the ITSS lessons. The adapted content structure was used to score main ideas, details, the structure's name, and signaling words (see Table 3).

All student input was parsed and checked against a custom dictionary and synonyms before being compared with the content structure for scoring. The parser was used to tag terms in the modified content structure tree as the keywords to be scored (see Table 3). For example, for main ideas, this essentially resulted in a list of main ideas for ITSS to score for a particular text. There was no overlap between keywords in the main idea and detail categories. The spell check was used to clean any spelling errors and apply the most appropriate replacement for a given term. The spell check was customized with spelling and typing errors produced by elementary and/or middle school students with these

Table 3
Scoring Recall for Lesson 14: Criteria, Example With Feedback, and Acceptable Response

Problem-and-solution top-level structure	
Idea	Type
Problem	Signaling
Listed	Main idea
Whales	Main idea
Endangered	Main idea
Extinction	Main idea
Possible	Detail
Species	Main idea
Seven	Main idea
Solution	Signaling
Declared	Main idea
Ocean	Main idea
Antarctic	Main idea
Sanctuary[a]	Main idea
Allowed	Detail
No	Detail
Hunting	Detail
1994	Detail
Commission	Main idea
Whaling	Main idea
International	Detail

Note. Paraphrases and misspellings were programmed into ITSS and counted as correct for each of the ideas in the above scoring structure. Criteria: Two out of two problem-and-solution signaling words to get positive (100%) feedback on signaling; six out of 12 main ideas to get positive (50%) feedback on main ideas; three out of six details to get positive (50%) feedback on details. Example with feedback: "Seven spieces of whales were on the endangered spiecies list. The solution to this possible extinction was saved by the International Whaling Commision. They did this by opening a whale sanctuary in the Antarctic Ocean" signaling = 1 or 50% (−), main ideas = 10 or 83% (+), and details = 2 or 33% (−) so I.T. gave accolades for main ideas and instructions for signaling and details. Student added "problem" signaling word and "1994" on second try: signaling =100% (+), main ideas = 83% (+), and details = 50%, so I.T. (Intelligent Tutor) gave positive feedback, "Wow, you are good at this!".
[a] Credited words or entire phrases counted as correct for *sanctuary* were sancuary, santchuary, santuary, sancsuary, sanchuary, sancutary, sentuary, sanctuay, sanctary, sanctolary, sanctry, sactionary, snctuary, sacuwaery, sansuwary, sencey, santuarie, sanctuuary, sancuwarry, sacnutary, sacuary, santuwary, sancuray, sanuitonary, santuuary, home, areas, not allowed to hunt, no hunting aloud, be free, safe, safety, harm free, haven, no hunting, no more hunting allowed, hunteers were not allowed, no whale hunting, no one can hunt there, no hunters allowed saf, havon, no hunting zone, hunting free zone, and no hunting allowed their.

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

texts in past studies. This data-cleaning step in ITSS was the correction of spelling errors and the removal of any extra characters produced by typing errors. A synonym table was constructed based on common synonyms for each word as well as paraphrases used by students in past studies. The synonym table was used to determine whether the students had used any synonyms for the keywords (see Note a in Table 3). After this data-cleaning process, the computer algorithm used benchmarks for main ideas, details, and signaling based on the content structure to match student input (i.e., typed text recalls).

A sample is shown in Table 3 for a text about whales from ITSS Lesson 14. The main ideas and details shown in Table 3 are based on the content structure. The automated ITSS scoring matched a student's input (i.e., Student 1 in Table 4) with the predefined trees for the whales text. In summary, ITSS scoring matches the student input to an expected set of keywords and their synonyms. Although we have explored more complicated scoring approaches, this simple approach is the scoring system used for this study.

This automated ITSS scoring approach was examined with 48 free recalls submitted by students in the individualized condition in response to two parallel texts from the fourth comparison lesson (whales or bears) with 43 main ideas and 50 details. Recalls were first independently scored by two trained human scorers and then compared with the automated ITSS scoring. The human scorers were an undergraduate student and the first author. The percentage agreement between these two scorers was 99% for main ideas and 97% for details. Percentage agreement between the undergraduate scorer and automated ITSS scoring was 95% for main ideas and 71% for details. Similarly, the percentage agreement between the first author and automated ITSS scoring was 97% for main ideas and 78% for details. The automated scoring appeared to be accurate and particularly so for main ideas. Such accuracy for main ideas was particularly critical because the individualization of instruction was based on the percentage of correct main ideas from the free recall tasks.

Feedback. I.T. provided elaborated feedback that involved scaffolding and good examples. Feedback events were the same for students in the two individualization conditions. Criterion for obtaining a correct response was 50% of the ideas in the structure for each kind of information scored (e.g., 50% main ideas in recall) and 100% of the signaling words (e.g., one comparison signaling word for the comparison structure and two for the problem-and solution structure: one for the problem part and one for the solution part). Criteria were determined by examining recall from fifth-grade students in prior studies (e.g., Meyer et al., 2002, 2010). For students not reaching the criterion for success on their first recall, three more trials, each with more scaffolding, were provided before moving the student to the next ITSS lesson.

I.T. provided feedback to the student, such as saying, "Your structure, main idea, and details are correct. Great job! But your signaling words were incorrect. Using the chart as your guide, rewrite the signaling words." For tasks focused on the generation of main ideas, students were also provided with additional help through pop-up windows providing model main ideas that could be viewed while students were correcting their main idea statements but could not simply be copied and pasted into students' answers. For example, after two unsuccessful attempts I.T. stated, "Please read my main idea and correct your work." Additionally, an example in Table 3 is provided for how feedback to students results from a combination of signaling, main idea, and detail scores computed by ITSS.

Differences between individualized and standard ITSS. In individualized ITSS, students received remediation lessons after evidence of poor performance or enrichment lessons after good performance (e.g., Figure 3), whereas lessons were not varied from a fixed sequence in standard ITSS. Students receiving remediation

Table 4
Individualization Criteria for Paths From First Problem-and Solution Recall to the Next Lesson

Criteria: Recall of main ideas in Lesson 14[a]	Sample of student recalls and main idea scores: % of main ideas	Resultant paths: Replacement texts for lesson 15[b]
Less than 25%	Student 1: "Wales are endagrd." Two main ideas (17%)	15ae for remediation: Three students (5%) traversed Path 14 to 15ae (again <25%) so then remediation Path 15ae (pig: alternate easy) to 15e (dog: easy)
Main idea recall >24% & <50%,	Student 2: "Whales are endangered spieces. There are only 7 different spieces and a solution to this problem is that they built a whale sanchuary for the whales to live in peace." Five main ideas (42%)	15: 32 students (49%) traversed Path 14 to 15; next 15 to 15a for remediation because recall on Lesson 15 was >24% & <50%
Main idea recall 50% or more	Student 3: "Seven spieces of whales were on the endangered spiecies list. The solution to this possible extinction was saved by the International Whaling Commision. They did this by opening a whale sanctuary in the Antarctic Ocean" 10 main ideas (83%)	15a for enrichment: 30 students (46%) Path 14 to 15a

Note. Individualized paths: Student 1 required the most remediation. Individualized Intelligent Tutoring of the Structure Strategy (ITSS) placed Student 1 on a remediation path with the easiest version of the 15th lesson, Lesson 15ae (see Figure 3a). Student 2 had some understanding of the structure strategy but was not ready for enrichment. ITSS placed Student 2 in Lesson 15; similar performance in Lesson 15 led to remediation via extra practice on the same lesson with the alternate text, Lesson 15a (see Figure 3b). ITSS placed Student 3 (main ideas = 83%) on an enrichment path from Lesson 14 to Lesson 15a because main ideas recalled in Lesson 14 were 50%+ (see Figure 3c & Table 3).
[a] See Table 3. [b] Replacement texts for the standard text in Lesson 15 were identified as alternate topic (a), easy version of standard text (e), and alternate topic easy (ae). Flesch-Kincaid readability grade levels are 4.6, 4.9, 5.6, and 7.9 for Lesson 15ae, Lesson 15e, Lesson 15, and Lesson 15a, respectively.

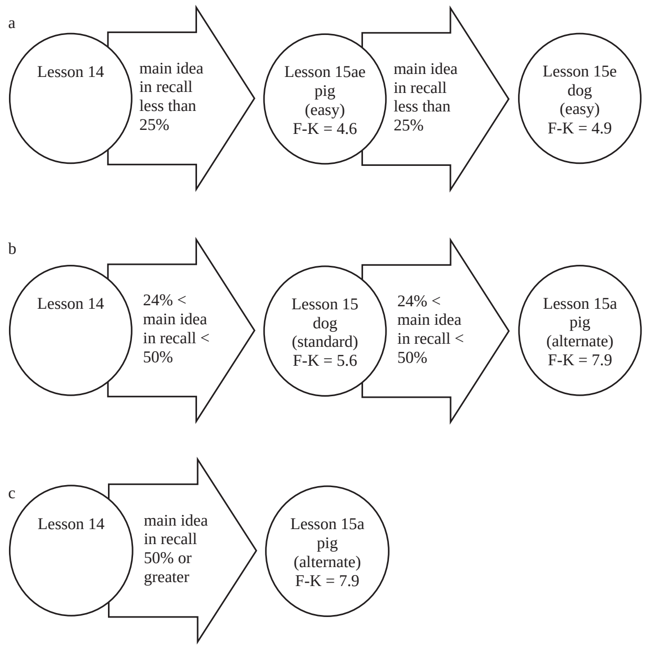

Figure 3. Three different paths from Lesson 14 to Lesson 15 taken by the 65 students in individualized Intelligent Tutoring of the Structure Strategy. (a) Remediation path with easy texts taken by three students scoring below 25%; (b) remediation path through standard and alternate texts taken by 32 students scoring from 25% to 49%; (c) enrichment path with alternate text taken by 30 students scoring at 50% or above. The path for all students in the standard condition is Lesson 14 to Lesson 15 (the same as first part of the path in Figure 3b), then all proceeding to the next standard lesson (Lesson 16). F-K = Flesch-Kincaid readability calculation for text.

worked on a parallel lesson with a new text of easier readability (see Figure 3a and Table 4: Student 1) or a similar readability level (see Figure 3b and Table 4: Student 2). Students receiving enrichment read the most difficult version of possible texts for the next lesson, usually the alternate topic text (see Figure 3c and Table 4: Student 3). Students with extremely high scores on the standardized reading comprehension test (higher than Grade 12) in the individually tailored condition were provided with a screen button to turn off the I.T.'s voice when reading texts in the lessons (see button in Figure 1).

Readability of texts. Currently, the most commonly used traditional readability measure is the Flesch-Kincaid grade level (Flesch, 1948; McNamara, Louwerse, Cai, & Graesser, 2005; Test Your Document's Readability, 2007). The measure ranges from U.S. grade-school levels of 0 to Grade 12 and is based on sentence length and number of syllables per word. We used the Flesch-Kincaid grade levels in the process of matching students to texts in the individualized ITSS condition. The average Flesch-Kincaid readability level for the texts in standard ITSS was 7.10 (SD = 2.37; range .5 to 12). If texts in standard ITSS had Flesch-Kincaid readability levels of fifth grade or higher, then easier versions of the texts were prepared at the third through fourth grade levels to help students in individualized ITSS needing remediation. Forty easier versions of texts in standard ITSS were prepared for use in the individualized condition. Students in individualized ITSS read easier versions with readability levels below the fifth grade level if they recalled less than 25% of the main ideas from the preceding lesson's text (e.g., Table 4; Figure 3a). Readability was reduced by (a) using more familiar words and (b) shortening sentences via deletion of specific details. We did not delete the logical structure among important ideas in our efforts to provide easier texts (see Meyer, 2003).

Alternate texts for the 29 practice lessons in standard ITSS provided further opportunity for individualizing instruction. Alternate texts were written with parallel structures and the same number of words, but different content. Alternate texts could be used for (a) remediation (repeating a preceding lesson's objectives with another text at the fifth-grade level or above) or (b) enrichment (i.e., usually skipping the original standard lesson and reading the alternate version or occasionally skipping a very easy lesson and/or working with both versions of a challenging lesson, such as comparing three unfamiliar ideas on at least four attributes with texts of seventh-grade readability). The alternate texts were primarily on less familiar topics and contained more multisyllabic words (compare columns 2 & 3 in Appendix A). On average the alternate texts had a significantly higher Flesch-Kincaid (F-K) grade level than their paired standard texts (29 alternate texts: M = 7.42, SD = 2.25; 29 paired standard texts: M = 7.07, SD = 2.48), paired samples $t(28) = 2.04$, $p = .026$. If students performed well on the preceding lesson with a main idea recall of 50% or more, then for enrichment they were sent to the alternate or standard lesson with the highest Flesch-Kincaid (F-K) readability level. Frequently, this was the alternate version (Figure 3c). For example, in standard ITSS, a comparison text involved comparing dogs and cats on response to owners, energy level, and bathroom habits (Flesch-Kincaid grade level = 3.4). The alternate text compared potbellied pigs and chinchillas on these same traits (Flesch-Kincaid grade level = 6.5).

Another example is shown in Tables 3 and 4 and Figure 3. This example refers to the 15th lesson in ITSS, which is the first practice lesson after the introduction of the problem and solution structure in Lessons 13 and 14 (see Appendix A). The text in Lesson 15 of standard ITSS presents problems with a dog and solutions to eliminate these troublesome behaviors. The alternate version of Lesson 15 is noted by *a* after the lesson number (i.e., Lesson 15a) in Table 4. The text for this lesson presents problems with a pig and solutions to eliminate these troublesome behaviors. As can be seen in Figure 3, the alternate text version in Lesson 15a had a higher Flesch-Kincaid grade level (Grade 7.9) than the standard version in Lesson 15 (Grade 5.6). Typical of the alternate texts, pigs are more novel pets than dogs. Figure 3c displays how Lesson 15a was used for enrichment (also see Student 3 in Table 4).

Easy versions also were written for the alternate lessons, and they are identified with *ae* (alternate topic easy) after the lesson number in Tables 3 and 4 and Figure 3. In the set of texts associated with the 15th lesson, Version 15ae had the easiest Flesch-Kincaid readability (Grade 4.6), followed by 15e (Grade 4.9), 15 (5.6), and 15a (7.9). Deleting the *potbellied* descriptor from the pig text as well as breaking up some lengthy descriptive sentences substantially reduced the readability level for the pig text. Students recalling less than 3 out of 12 main ideas in Lesson 14 were sent on a remediation path to Lesson 15ae, the easiest text of the four possible texts for Lesson 15. As shown in Figure 3a, three students scored below 25% on both Lesson 14 and Lesson

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

15ae; individualized ITSS next sent these three students to Lesson 15e for further remediation.

Paths from one lesson to the next. Based on a student's performance in remembering the main ideas of an expository text in the recall task, individualized ITSS placed the student into an appropriate next lesson rather than followed the standard ITSS sequence of lessons (e.g., Tables 3 and 4; Figure 3; Appendix A). Table 3 displays an example of a scoring structure used in ITSS to score recall. Also specified in the table is an example of acceptable paraphrases and misspellings in ITSS for scoring a main idea. Tables 3 and 4 focus on main ideas produced by students in Lesson 14, a modeled lesson in the first problem-and-solution lessons. Table 4 also provides examples and scoring of three students' recalls in Lesson 14 that led to different paths traversed in the more individually tailored condition. Table 4 and Figure 3 display the paths taken by the 65 students in the individualized condition from Lesson 14 through the set of lessons associated with the 15th lesson.

There were 200 possible paths between lessons for individualized ITSS. These 200 potential paths included those going between two standard lessons as well as between all potential combinations of standard, remediation, and enrichment lessons. Of these potential paths 154 were taken by at least one student in the individualized ITSS condition.

Remediation and enrichment paths between one lesson and the next were identified for each of the 65 students in the individualized ITSS condition. Identified were (a) 19 paths moving students to an easy version of a standard ITSS lesson, (b) nine paths providing students with repetition of the same lesson but with another text with a readability level above fourth grade, (c) 14 paths providing individualized students with repetition of the same lesson but with an easier alternative text, and (d) 13 enrichment paths moving students to a more difficult version of a standard ITSS lesson.

The lessons traversed by each student throughout individualized ITSS instruction were categorized into four paths: all remediation lessons (25%), a balance of remediation and enrichment lessons (34%), mainly enrichment lessons (20%), and all enrichment lessons (21%). Table 5 displays these paths for the three reading ability groups and, not surprisingly, shows that better readers received more enrichment lessons, whereas poorer readers received more remediation lessons.

A student in the individualized condition could potentially move more slowly through ITSS lessons (e.g., 5 paths: Lesson 3 to 4e to 4ae to 5e to 5ae to 6) than students in the standard ITSS (i.e., 3 paths: Lesson 3 to 4 to 5 to 6; see first column in Appendix A). ITSS recorded the number of the lesson and the page on which each student was working on the last day they worked with ITSS. An independent t test was run to examine whether students in the two ITSS conditions varied in how far they progressed through the lessons. Students in the two conditions did not vary significantly on the number of the last lesson completed in ITSS (individualized ITSS: $M = 35.14$, $SD = 13.64$; standard ITSS: $M = 37.45$, $SD = 17.55$), $t(129) = 0.84$, $p = .404$. These findings are similar to those shown in Table 2 for the amount of time spent working in ITSS, number of texts read, and number of lessons worked on by students in the two conditions.

Thus, students did not complete more lessons per structure in the individualized condition than in the standard condition. However, the difficulty of the texts in the lessons better matched each reader's online needs and performance. For example, there are 12 comparison lessons in standard ITSS. Eighty percent of the students in individualized ITSS also completed a total of 12 comparison lessons. More specifically, the 13 students receiving just enrichment lessons (see Table 5) completed 12 comparison lessons; however, the practice texts for these students were more difficult and/or less familiar than were those of the students in standard ITSS. Most (80%) of the 15 students receiving mainly enrichment lessons also completed 12 comparison lessons, but three students completed a total of 15 comparison lessons. Additionally, most (86%) of the students receiving a balance of enrichment and remediation completed a total of 12 comparison lessons. For the group receiving all remediation lessons (see Table 5), 53% of the students completed a total of 12 comparison lessons with the remainder completing up to 15 comparison lessons.

Table 5
Classification of Differentiation Paths for 65 Individualized Students Varying in Reading Comprehension

Classification of differentiation paths in individualized ITSS	Reading comprehension ability groups: Low	Middle	High
All remediation	7	7	1
Balance of remediation & enrichment	8	6	8
Mainly enrichment	2	4	9
All enrichment	2	7	4

Note. $\chi^2(6, N = 65) = 12.85$, $p = .045$. ITSS = Intelligent Tutoring of the Structure Strategy.

Data Sources, Variables, and Reliabilities

Standardized reading comprehension test. The GSRT (Wiederholt & Blalock, 2000) is a multiple choice reading comprehension test that allows for group administration with reliable alternative forms and tests deep comprehension processes that include finding the main idea and reasoning with the text's main idea. Most of the 13 passages in the test are short narratives arranged in difficulty from extremely easy to complex. The GSRT is designed to test all readers 7 years through 25 years of age with the same test. Average alternate-form reliability was reported in the test manual at .85 (i.e., .87 for 10-year-olds), and delayed alternate-form reliability was reported at .83. Coefficient alpha reported for Forms A and B were .95 and .94, respectively.

The text structures presented by ITSS are represented by at least one passage in the GSRT as a major organizing structure within an overall narrative genre. All of the multiple-choice questions (5 per text) in the GSRT require at least paraphrases of text information. The GSRT questions often required construction of an accurate main idea or reasoning with one (P. A. Alexander, personal communication, July 15, 2008). The GSRT focused on one-paragraph narrative texts and multiple-choice questions, whereas ITSS lessons focused on multiple-paragraph expository texts of varying lengths and free recall. Both involved main ideas and relatively short texts.

Due to concern about the use of grade-equivalent scoring and interpretations (e.g., Reynolds & Kamphaus, 2003; Schulz & Nice-

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

wander, 1997), raw scores were used in the current study. Cronbach's alpha coefficients (Cronbach, 1951) of .95 indicated strong internal consistency for the two forms in a pilot sample and the current sample. Additionally, factor analyses were run for the two forms, and the factor structures were comparable. There were no statistically significant form effects on raw test scores. Experimental conditions in the study were examined with repeated measures analysis on GSRT raw scores.

Experimenter-designed measures of reading comprehension. A set of three equivalent passages was prepared with the comparison structure. Each comparison text had 128 words, 15 sentences, 96 idea units, and equivalent readability scores (Flesch-Kincaid grade level = 6.7). Each text compared two types of turtles, monkeys, or penguins on a number of attributes (see Appendix B: Adelie versus Emperor penguins). The problem-and-solution set of three equivalent passages each had 98 words, 72 idea units, and the same readability scores (Flesch-Kincaid grade level = 10.8). The texts within each set had equivalent scores on traditional measures of readability and aspects of text cohesion and/or coherence (see Meyer, 2003). Each text presented a relatively unfamiliar problem, its cause, and a solution that eliminated the cause of the problem on the topics of cats, dogs, or rats. The article about rats came from a newspaper. The rats and penguins texts (Form RP) are displayed in Appendix B.

For the recall task, students were asked to write all they could remember after reading each text and after placing the text out of sight in an envelope. Total recall and TLS were measured from the recall protocols. The interrater reliability coefficient for 10% of the total recall scores was .96. Interrater reliability for TLS scores also was high; the reliability coefficient was .98. The TLS scores are an indication of how well students organized their understanding of the text and used the recall pattern for a structure (e.g., Figure 2b). Percentage of agreement between scorers is discussed later under scoring.

Signaling test. The signaling test (see penguins example in Appendix B) involved filling in missing signaling words in a two-paragraph comparison text. Students completed the signaling test by filling in four comparative signaling words. Students reread the passage prior to recalling it. The interrater reliability coefficient for the signaling test was .98. Reasons for only testing comparison signaling words on the signaling test included the 100% completion rate of lessons focusing on comparison signaling words and time constraints for testing within the schools.

The signaling test was designed to examine whether we could develop a quick, reliable assessment of text structure knowledge for future ITSS assessments and classroom teachers. To investigate classroom applications, a multiparagraph text from a fifth-grade social studies textbook was administered prior to starting ITSS. Blanks were substituted for three comparative signaling words and two problem-and-solution signaling words. Correlations between the signaling test and the classroom application to a social studies text as well as to TLS (top-level structure) scores for recall of the comparison texts are displayed in Table 6. The data suggest some potential for the signaling test as a quick, practical assessment of structure strategy use in the classroom and for ITSS scoring.

Achievements goals. The specific aim of goal assessment was to determine students' goal orientation while working with ITSS. Three scales from the Patterns of Adaptive Learning Study (PALS; Midgley et al., 2000) were used to measure goals via computer on the last ITSS computer class. The PALS was selected for three primary reasons. First, it has been widely used in related research (e.g., Anderman, 1999). Second, it has been validated in studies with learners from similar populations as those in our study. Third, the psychometric properties of the PALS scales are reported to be stronger in several aspects when compared with other existing goal orientation instruments (see Day, Radosevich, & Chasteen, 2003).

Table 6
Correlations Between the Signaling Test (Comparison Signaling Words) and Other Measures

Measure	Signaling test: Pretest	Signaling test: Immediate posttest	Signaling test: Posttest month after ITSS
Pretest social studies text application	$.42^{*}$	$.32^{*}$	$.38^{*}$
Pretest TLS score	$.54^{*}$	$.43^{*}$	$.49^{*}$
TLS score immediately after ITSS	$.39^{*}$	$.60^{*}$	$.59^{*}$
TLS score 1 month after ITSS	$.30^{*}$	$.59^{*}$	$.60^{*}$

Note. ITSS = Intelligent Tutoring of the Structure Strategy; TLS = top-level structure.
* $p < .0005$.

Learning goals. To measure learning goal orientation, the Mastery Goal Orientation Revised Scale from the PALS was used. The scale has a reported coefficient alpha of .85 (Midgley et al., 2000). All five items from the scale were used. Only slight modifications were made to items to provide a context for the structure strategy work instead of school in general. For example, the original PALS item was "*One of my goals in this class is to learn as much as I can*." We modified the item to state, "*One of my goals when I do this work is to learn as much as I can*." Students were instructed to rate on a 5-point scale their agreement with the statements in terms of what they thought about working with ITSS. Scores on the measure range from 5 to 25, and Cronbach's alpha reliability (Cronbach, 1951) for our sample was .91.

Performance goals. The Performance-Approach Goal Orientation Revised Scale from the PALS was used to measure learners' performance approach tendencies. The Performance-Approach Goal Orientation Revised Scale consists of five items with a reported scale reliability of .89. All items were used and were only slightly modified to provide context for the structure strategy. An example item is "*One of my goals is to look smart in comparison to the other students.*" Scores on the measure range from 5 to 25, and Cronbach's alpha reliability was .87 for our sample. We also included all of the items from the PALS Performance-Avoidance Goal Orientation Revised Scale. These four items have a reported scale reliability of .74. An example item is "*One of the goals is to avoid looking like I have trouble doing this work*." Scores on the measure range from 4 to 20, and Cronbach's alpha was .87 for our sample.

Questionnaires. A computer attitudes questionnaire and self-efficacy questionnaire were administered via the computer. Each was administered twice: at the beginning of the first ITSS session and at the end of the last ITSS session.

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

Computer attitudes. We used the same computer attitudes questionnaire as administered to fifth-grade students in Meyer et al. (2002). Participants used a 5-point Likert scale to answer 21 questions about attitudes toward computers (5 = *strong agreement* with positive statements about computer-based technology, 4 = *somewhat agreement*, 3 = *indifference*; 2 = *somewhat disagreement*, 1 = *strong disagreement*). Both positive and negative statements about computers occurred on the questionnaire, and the negative statements were reversed scored before their addition to the total score on the questionnaire. Internal consistency, measured by Cronbach's alpha (Cronbach, 1951), was reported at .64 for pretest and .79 for posttest administration. Krauss and Hoyer (1984) designed the questionnaire to measure attitudes toward computer-based technology. Cronbach's alpha statistics for our sample were .83 on the pretest and .86 on the posttest.

Self-efficacy. To measure self-efficacy, a 23-item (4-point Likert scale; reported reliability = .86) questionnaire was administered (Sherer, Maddux, Mercandante, Prentice-Dunn, Jacobs, & Rogers, 1982). Older adult tutors and fifth-grade tutees showed increased self-efficacy on this measure after structure strategy instruction (Meyer et al., 2002). Cronbach's alpha statistics for this measure with the current sample was .87 on the pretest and .85 on the posttest.

Procedure

In September, students were randomly assigned to take either Form A or Form B of the GSRT. The testing was conducted in a large auditorium, lasted 2.5 hr, and was monitored by classroom teachers and teacher aides and two members of the research team. Forms A and B of the GSRT were counterbalanced over the pretest and the immediate posttest (Posttest 1) that was administered after the conclusion of ITSS.

Students were stratified by reading comprehension abilities and elementary school. Then, using a stratified random assignment procedure, students were randomly assigned to the six conditions in the experiment (two design variations—standard ITSS versus individually tailored ITSS × 3 experimenter designed test forms—Form CT = cats and turtles, DM = dogs and monkeys, or RP = rats and penguins). The three forms of the experimenter-designed materials were counterbalanced over the three times of testing (pretest, immediate posttest, and 1-month delayed posttest). Students eventually received all three forms, but in different orders. A third of the students were assigned through stratified random assignment to Form CT on the pretest, DM on the immediate posttest, and RP on the delayed posttest. Another third took Form DM on the pretest, RP on the immediate posttest, and CT on the delayed posttest. The final third took form RP, CT, and DM on the pretest, immediate posttest, and delayed posttest, respectively.

Prior to using the ITSS program, all students were given the researcher-designed pretests. Students were brought together in the large auditorium again with the same monitoring conditions as their GSRT testing and completed the tests in 50 min.

Students at each elementary school used ITSS three times a week for 30 min. ITSS replaced 90 min a week of regularly scheduled social studies for all students in both schools. Data from students without parental consent were excluded from the study. During the first session with the computer, students were familiarized with ITSS procedures, and students took the computer attitudes and self-efficacy questionnaires. At the end of the last ITSS session in April, students again took the computer attitudes and self-efficacy questionnaires as well as the goal orientation questions about their work in ITSS.

Under the same testing conditions as provided for the pretests, students completed at posttest the GSRT (A or B, taking the alternative of what they completed at the pretest) and the appropriate researcher-designed counterbalanced test (Form CT, DM, or RP). In April after ITSS, students were given the immediate posttest (one session for the GSRT and another session for the researcher-designed counterbalanced test). A month later in May, students were given the delayed posttest (Posttest 2), which involved only the researcher-designed counterbalanced test.

Scoring

The scorers for all measures were blind to the experimental condition of the participants. The prose analysis system of Meyer (1975, 1985) was used to score the experimenter-designed measures. Scoring manuals based on Meyer's approach to discourse analysis were prepared for each passage and scoring structures were typed into an adapted Excel program to score and automatically tally idea units from the texts and the interrelations among these idea units. A graduate student in school psychology with a prior year of mentored training in the scoring procedure scored all of the free recall data. At least 10% of the data from each of the measures were randomly selected from the two conditions, three times of testing, and three forms and scored by an experienced researcher in educational psychology. Percentage agreement between scorers for total recall scores for the problem-and-solution set of texts was 92%; percentage agreement between scorers for total recall scores for the comparison set of texts was 91%. Sample recalls and scores for total recall and TLS can be found in Appendix B along with the texts from form RP.

TLS. The TLS scale (Meyer et al., 1980, 2002, 1989, 2010) was used to appraise the similarity between the organization of a student's recall protocol and the text structure organizing the article read by the student. As seen in Table 7, the scale runs from 1 (*no correspondence*) to 9 (*explicit match*). TLS scores of 6 or greater (Meyer, 1985) indicated use of the structure strategy on a particular text. For the problem-and-solution texts, scores of 6 to 9 indicated that all the problem information was presented together followed by all the solution information. To receive scores of 6 to 9, the order of the recall in terms of problem(s) and solution(s) had to match that of the text, but the content of the problems or solutions did not have to match that of the text. A score of 7 indicated that in addition, the participant used a signaling word to explicitly identify a problem. A score of 8 indicated that a solution had been signaled, and a score of 9 showed that both the problem and solution parts of the problem-and-solution structure had been explicitly signaled. Percentage agreement between scorers on TLS of the problem-and-solution texts was 97%.

For the comparison texts, scores of 6 to 9 indicated that the recall was organized into two distinct, contrasting parts. A score of 7 indicated that the participant also explicitly signaled the first idea compared with a comparison signaling word (e.g., "Leatherback turtles and hawksbill turtles have *differences*.

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

Table 7
Top-Level Structure Scale

Score	Description of scale (examples & clarifications)
1	No correspondence. (“I don’t know,” “I don’t remember,” or 90% or less of ideas from the article in a bizarre recall.)
2	Descriptive list of ideas about the text with no indication in any sentences about text structure used as the top-level structure of the article. (“It’s about penguins.” “Psychologists are working with rats.”)
3	More than a descriptive list (use of other structures, such as cause and effect for a comparison article, and no ideas organized with the article’s comparison structure).
4	Like 2 above, but one of the listed descriptions is organized like the article’s overall structure. (The student’s recall is organized like a list of things remembered.)
5	Like 3 above, but within a sentence or two adjacent sentences student expresses the same text structure as that used to organize the article.
6	Top-level structure of recall matches that of the article, but no explicit signaling.
7	Top-level structure of recall matches that of the article and explicit signaling of first part of the text structure.
8	Top-level structure of recall matches that of the article and explicit signaling of second part of the text structure.
9	Top-level structure of recall matches that of the article and explicit signaling of both parts of text structure.

Note. Adapted from “Web-Based Tutoring of the Structure Strategy With or Without Elaborated Feedback or Choice for Fifth- and Seventh-Grade Readers,” by B. J. F. Meyer, K. Wijekumar, W. Middlemiss, K. Higley, P. Lei, C. Meier, and J. Spielvogel, 2010, *Reading Research Quarterly*, 45, p. 80. Copyright 2010 by the International Reading Association. Adapted with permission.

Leatherbacks can get to be 8 ft long and . . .”). A score of 8 indicated that the recall was organized in contrasting parts with a least one comparison signaling word explicitly cuing the second idea compared (e.g., “Emperor penguins are 4 feet tall and 90 pounds. They eat seafood. *In contrast.* Adalie penguins only get about 2 feet tall and weight 11 pounds. Their diet is krill.”). A score of 9 indicated that the recall was organized in contrasting parts, and each part was explicitly signaled. Percentage agreement between scorers on TLS of the comparison texts was 96%.

Signaling test. The signaling test was scored on a 7-point scale for each of four missing signaling words in the comparison main idea task only. Scores per missing signaling words ranged from 7 points for verbatim use of the intended signaling word (e.g., however) to 1 point for use of a content word (e.g., soldier) that made little semantic or grammatical sense. A trained educational psychology graduate student scored the signaling test; a stratified random sample of 10% of the data from the signaling test was scored by another trained student. Percentage agreement was 97%.

Design

Pretest–posttest design with testing forms counterbalanced over testing time. A pretest, an immediate posttest after ITSS training, and a 1-month delayed posttest design were used to compare two methods of structure strategy intervention delivered via Internet with computer-based tutors. Students were randomly assigned to the more individually tailored ITSS versus the standard ITSS instruction. Individualized ITSS differentiated the sequence, difficulty, and amount of practice to meet students’ online performance needs. Counterbalanced forms over testing times increased the rigor of the design. There were no significant form effects on the experimenter-designed reading measures related to the comparison texts: total recall scores, TLS, and the signaling test (Wilks’s Λ = .98), $F(6, 252) = 0.48$, $p = .823$; this finding corroborates past research on these texts (Meyer et al., 2010). However, in the earlier study for the problem-and-solution set, the cats form was found to be more difficult than the rats and dogs forms for total recall, but not TLS. With the current sample there was not an order effect for either total recall scores or TLS (Wilks’s Λ = .90), $F(4, 252) = 0.90$, $p = .465$, but the pattern of poorer total recall scores on cats was found for each testing time, reaching significance on the pretest. As a precaution and for compatibility in scores with the earlier study, linear equating was used to adjust form difference in difficulty (Kolen & Brennan, 1995, p. 30). Equated raw scores for total recall were then converted to standard scores (*T* scores) for each problem-and-solution passage standardized across the three times of testing so that gain scores over time would not be removed by the standardization. Test form (CT, DM, or RP) will not be included in subsequent analyses because form effects were either negligible or resolved through linear equating, and forms of the measures were counterbalanced over the three times of testing.

Random assignment of test forms and instruction conditions. A simple random assignment procedure was used to randomly assign students to two forms of the standardized reading comprehension test that were counterbalanced over the pretest and immediate posttest. Stratified random assignment was used to assign students to an individualization condition and the three forms of the experimenter-designed materials. Stratified random assignment was based on the student’s composite reading score. This score was calculated for each student based on GSRT scores and reading assessment scores provided by the school district. Pretest *z* scores from the GSRT constituted half of the composite reading measure. The other half of the composite score was an average of *z* scores from the current classroom teachers’ ratings of their students’ reading comprehension and reading scores from the prior school year on the *Stanford Achievement Test* (SAT) and the Pennsylvania state assessment.

There were 11 strata in one school and 12 in the other school, with six students per stratum (e.g., the 6 most outstanding readers per school, the 6 next best readers, etc.). Students within each stratum were randomly assigned to one of six conditions. The conditions were individualization of ITSS (individualized ITSS or standard ITSS) and three forms of experimenter-designed measures of reading comprehension that were counterbalanced over the three times of testing (Forms CT, DM, & RP). The top six readers in a school were randomly assigned to one of the six experimental conditions with a random number table. The same procedure was followed for the next six readers on the ranked list. The procedure was repeated until all students were assigned to experimental conditions and testing forms. Seven students were dropped from the data set due to relocation prior to posttest, three students were dropped by request of the parents to withhold the data from the study, and two students were dropped because they completed few ITSS lessons due to absences related to special services or illness.

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

Missing data. Experimenter-designed pretests and GSRT pretest and posttest were complete with no missing data. As displayed in Appendix C, there were a few missing data entries for some of the tasks on the experimenter-designed posttests (Posttest 1 immediately after ITSS instruction and Posttest 2 a month after ITSS instruction) and up to 10 missing cases for the various questionnaires. The number of missing cases ranged from 0.8% (experimenter-designed measures at posttest 1) to 7.6% (posttest achievement goal questionnaire).

As is noted in Appendix C, no participants missed all of the measures collected after ITSS. Only one student missed the experimenter-designed recall measures on the immediate posttest (Posttest 1). Also, one other participant missed the experimenter-designed recall measures on the month-delayed posttest (Posttest 2; see frequencies for signaling test and problem-and-solution recall in Appendix C).

Other participants with some missing data had only partially missing data on experimenter-designed recall measures. The problem and solution text recall came first in the packet of tasks, and all students who attended the testing sessions completed the problem-and-solution recall task. The problem-and-solution recall task was followed by the signaling test and followed by recall of the comparison text. One student skipped the signaling test but completed all of the other tasks. The other students (three in Posttest 1 and two in Posttest 2) did not finish the comparison recall, the last task.

The posttest questionnaires were collected during the last computer lab of regularly scheduled ITSS instruction. Five students with missing data on all questionnaires were absent on the last day in the computer lab. The achievement goals questions came near the end of the testing session, and five students did not progress through the achievement goals questions.

Occurrence of some missing values was not completely random, so an EM (expectation-maximization) estimation method was used to impute missing values (Peng, Harwell, Liou, & Ehman, 2007). We used the EM estimation method provided by the missing value analysis available in SPSS (2007). Quantitative variables (recall and main idea pretest scores, GSRT pre- and posttest scores) and categorical variables (three reading ability groups based on the GSRT pretest, individualization condition, and test form) were used to impute missing values with the EM estimation method.

The major analysis concerning the findings from the standardized reading comprehension test had no missing data, so the missing data issue had no effect on these findings. The other analyses involving missing data (i.e., recall, signaling test, achievement goals, computer attitudes, self-efficacy) were conducted with and without missing data. The findings were the same regarding the pattern of results and statistical significance for main effects and interactions, regardless of the use of imputed missing data scores.

Results

The primary research question was whether the design variation (individualized ITSS vs. standard ITSS) affected reading comprehension as measured by experimenter-designed tests and a standardized reading comprehension test. The goal of the analysis of performance on the standardized reading test was to determine whether the design variation affected transfer to a standardized reading comprehension test. Experimenter-designed recall measures were intended to be similar to those used in ITSS but contained different reading passages. During each testing session students were requested to recall a text with a comparison TLS and another text with a problem-and-solution TLS (with an embedded cause-and-effect in the problem). In our sample, 88% of the students completed the lessons with instruction about these two text structures. There were no significant differences between the individualized conditions in whether students had completed instruction focused on these two text structures, $\chi^2(1, N = 131) = .001$, $p = .974$.

All statistical tests in this study were assessed with an alpha level of $p < .05$. We report obtained p values for all analyses (except those of .000, which are reported as $p < .0005$).

Did Greater Individualization Affect Performance on Experimenter-Designed Measures?

Recall and TLS. A repeated-measures multivariate analysis of variance (MANOVA) was conducted to examine total recall scores and TLS scores assessed on the recall tasks for the comparison and problem-and-solution texts. The repeated measure in the MANOVA was time of testing (pretest, immediate posttest, and delayed posttest). Between-groups factors were design features varying on individualized instruction (more individually tailored vs. standard ITSS) and reading ability (high, middle, low). Averages and standard deviations for total recall scores and TLS scores are shown in Table 8. The only statistically significant main effects from the MANOVA were time of testing (Wilks's $\Lambda = .34$), $F(8, 118) = 28.35$, $p < .0005$, and reading ability (Wilks's $\Lambda = .77$), $F(8, 246) = 4.09$, $p < .0005$. Students remembered more after ITSS instruction than before ITSS instruction (see follow-up analyses for time of testing effect in Table 9). The factor of reading ability was statistically significant for all measures at each time of testing. Contrary to predictions, the individualization condition (individualized or standard) did not affect pretest to posttest gains (interaction between design condition & time of testing: Wilks's $\Lambda = .91$), $F(8, 118) = 1.45$, $p = .183$. None of the other interactions were statistically significant: Condition $\times$ Ability: Wilks's $\Lambda = .95$, $F(8, 244) = 0.77$, $p = .632$; Condition $\times$ Time: Wilks's $\Lambda = .91$, $F(8, 118) = 1.49$, $p = .169$; Time $\times$ Ability: Wilks's $\Lambda = .88$, $F(16, 236) = 96$, $p = .503$; and three-factor interaction: Wilks's $\Lambda = .90$, $F(16, 236) = 0.81$, $p = .670$.

As seen in Table 9, pretest performance was significantly lower than performance on the immediate posttest (effect sizes $d = 0.75$ to 1.45)[2] and delayed posttest ($d = 0.63$ to 1.43). Regardless of design feature variation, students improved the amount of information they could remember from text and the organization of their recalls in terms of the TLS. Both of these skills, recall and organization, were explicitly and repeatedly taught in both individualized and standard ITSS. Additionally, examinations of differences between immediate and delayed posttests suggest maintenance overall of the instruction a month after completion.

[2] Throughout the report, effect size was measured by standardized difference, $d =$ (Mean 1$-$ Mean 2)/SD, where SD was the standard deviation on the pretest (for pretest and posttest difference) or the standard deviation for the control group (for experimental and control group difference).

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

Table 8
Individualization Condition and Reading Level Means for Total Recall and TLS Over Time on Two Text Types

Recall variable condition (n in cell)	Comparison recall mean raw scores (*SD*)			Problem and solution recall mean T scores (*SD*)		
	Pretest	Posttest 1	Posttest 2	Pretest	Posttest 1	Posttest 2
High ability readers						
Total recall						
Standard (n = 22)	25.09 (17.85)	43.68 (20.78)	38.94 (19.59)	46.26 (7.21)	56.53 (10.63)	55.32 (10.81)
Individualized (n = 22)	27.00 (13.82)	47.77 (21.43)	46.50 (24.32)	48.66 (6.98)	57.77 (9.37)	55.61 (9.71)
Top-level structure (TLS)						
Standard (n = 22)	5.64 (2.38)	7.41 (1.74)	7.23 (2.02)	3.36 (2.08)	6.41 (2.48)	6.25 (3.19)
Individualized (n = 22)	5.18 (2.23)	7.27 (2.30)	6.91 (2.56)	4.77 (2.56)	7.36 (1.89)	5.73 (2.96)
Middle ability readers						
Total recall						
Standard (n = 20)	24.85 (17.05)	39.63 (20.43)	33.24 (21.25)	45.45 (6.93)	53.76 (10.38)	54.34 (9.44)
Individualized (n = 24)	25.25 (17.91)	40.54 (21.03)	32.54 (23.00)	44.06 (6.14)	53.99 (9.36)	52.33 (9.95)
Top-level structure (TLS)						
Standard (n = 20)	5.35 (2.37)	6.84 (2.25)	6.91 (2.60)	2.75 (1.71)	5.75 (2.79)	5.85 (2.70)
Individualized (n = 24)	5.04 (1.90)	7.00 (1.99)	6.38 (2.48)	2.58 (1.32)	5.70 (2.75)	4.58 (2.90)
Lower ability readers						
Total recall						
Standard (n = 24)	15.62 (14.24)	29.24 (17.37)	26.30 (20.07)	42.75 (4.99)	49.66 (9.98)	44.63 (10.75)
Individualized (n = 19)	15.00 (8.90)	29.89 (18.71)	22.58 (17.61)	41.44 (5.21)	48.56 (6.87)	48.98 (9.17)
Top-level structure (TLS)						
Standard (n = 24)	3.92 (1.84)	5.76 (2.53)	5.17 (2.53)	2.29 (1.33)	4.04 (2.37)	4.00 (2.74)
Individualized (n = 19)	4.05 (1.75)	5.37 (2.43)	5.00 (2.94)	2.47 (1.54)	4.89 (2.87)	4.05 (2.65)

Note. Comparison recall raw scores $N = 131$. Problem and solution recall T scores $N = 131$.

Specifically, differences were not statistically significant between the two posttests for five of the eight comparisons in Table 9.

In addition, effect sizes were minimal for seven of the eight comparisons in Table 9 between performance on immediate and delayed posttests ($d = -0.01$ to -0.26). For one of the eight comparisons, the effect size was not minimal (−.47; last row and column of Table 9). Students in the individualized condition dropped from an average of 6.07 on TLS for the problem-and-solution text immediately after ITSS to an average of 4.82 on TLS for the same structure after a month without ITSS. As seen in

Table 9
Means (Standard Deviations) and Effect Sizes for Recall and TLS Over the Three Testing Sessions for Individualization Conditions

Individualizing condition[b]	*M (SD)*			d^a		
	Pretest	Posttest 1	Posttest 2	Posttest 1–Pretest	Posttest 2–Pretest	Posttest 2–Posttest 1
Comparison total recall						
Standard	21.58 (16.74)	37.20 (20.19)	32.61 (20.66)	0.93^{***}	0.66^{***}	-0.23^{*c}
Individualized	22.85 (15.05)	39.88 (21.43)	34.35 (23.76)	1.13^{***}	0.76^{***}	-0.26^{**}
Problem-and-solution total recall						
Standard	44.74 (6.48)	53.19 (10.57)	51.14 (11.38)	1.30^{***}	0.99^{***}	−0.19
Individualized	44.85 (6.77)	53.68 (9.47)	52.46 (9.86)	1.30^{***}	1.12^{***}	−0.13
Comparison top-level structure						
Standard	4.92 (2.30)	6.64 (2.29)	6.38 (2.54)	0.75^{***}	0.63^{***}	−0.11
Individualized	4.80 (2.01)	6.62 (2.34)	6.15 (2.72)	0.91^{***}	0.67^{***}	−0.20
Problem-and-solution top-level structure						
Standard	2.79 (1.76)	5.35 (2.70)	5.31 (3.01)	1.45^{***}	1.43^{***}	−0.01
Individualized	3.29 (2.14)	6.07 (2.67)	4.82 (2.89)	1.30^{***}	0.71^{***}	-0.47^{***}

[a] Difference between pretest and Posttest 1 or Posttest 2 divided by the standard deviation on the pretest and difference between Posttest 2 and Posttest 1 divided by the standard deviation on Posttest 1. [b] Standard Intelligent Tutoring of the Structure Strategy (ITSS) $n = 66$ for all cells; individualized ITSS $n = 65$ for all cells. [c] Dependent t tests between two testing times (i.e., Posttest 1 − Pretest).
* Statistically significant at $p < .05$. ** Significant at $p < .005$. *** Dependent t tests statistically significant at $p < .0005$.

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

Table 9, this is a significant drop in performance. In contrast, with the same measure, students in the standard group did not experience a significant drop in performance ($M = 5.35$ immediately and $M = 5.31$ a month later; see Table 9). The two conditions did not vary significantly on problem-and-solution TLS prior to instruction, $t(123.58) = 1.47$, $p = .143$. At the immediate posttest (see Table 9), students in the individualized condition tended to outperform students in the standard condition, $t(129) = 1.54$, $p = .064$. The trend reversed, but not significantly on the delayed posttest, $t(129) = 0.96$, $p = .341$. The problem-and-solution structure is one of the most difficult to use effectively with the structure strategy (e.g., Meyer et al., 2010; note in Table 9: higher TLS scores and no significant decline a month after ITSS with the comparison structure.) The pattern of results tends to show initial superiority in learning the problem-and-solution structure by students in the individualized condition over students in the standard condition, but a month after completion of ITSS, the two groups performed at similar levels. However, as indicated in the reported results for the repeated-measures MANOVA, the interaction between time of testing and individualization condition was not statistically significant. Tests of within-subjects contrasts for the interaction approached significance for the TLS scores on the problem-and-solution text; quadratic: $F(1, 125) = 2.85$, $MSE = 9.66$, $p = .094$; linear: $F(1, 125) = 4.36$, $MSE = 18.00$, $p = .039$.

Some evidence for the greater difficulty in the problem-and-solution TLS is seen in Table 10, which follows up the significant main effect for reading ability. TLS scores of 6 or greater indicate that the recall pattern (i.e., Figure 2b) for a particular text structure was used to organize the recall. Overall (see Table 10), this was more prevalent for the comparison structure than the problem-and-solution structure. For the comparison TLS, Tukey's honestly significant difference (HSD) multiple comparisons tests showed statistically significant differences between low ability readers and both middle ($p = .002$) and high ability readers ($p < .0005$). The difference between middle and high ability readers was not statistically significant ($p = .623$). The pattern for the problem-and-solution TLS was different. Statistically significant differences were found between low ability readers and high ability readers ($p < .0005$) and also between middle and high ability readers ($p = .015$). The p value was .057 between low and middle ability readers. Patterns were the same among groups for recall on both passages. High ability readers recalled more information than did low ability readers ($p < .0005$). Middle ability readers recalled more information than did low ability readers ($p = .017$ and .008 for the comparison and problem-and-solution texts, respectively). High and middle ability readers did not differ ($p = .244$ and $p = .145$ for the comparison and problem-and-solution texts, respectively).

In summary, the predicted time by condition interaction on the experimenter-designed recall measures was not found. Clearly, students in both groups made similar gains in recall after both versions of ITSS. Similar findings were apparent for structure strategy use (TLS scores) on the comparison text. For structure strategy use with the more difficult problem-and-solution (with embedded major cause), there was some marginal support for students in the individualized condition acquiring better understanding immediate after ITSS than students in standard ITSS. However, any evidence for superior understanding by the individualized group over the standard group disappeared without continued ITSS instruction.

Signaling test. An analysis of variance (ANOVA) with repeated measures was used to examine scores on the signaling test. Three factors were entered as predictor variables: individualization condition, reading ability, and time of testing with repeated measures on time of testing. Descriptive statistics for the cells can be found in Table 11. Two main effects and one interaction were statistically significant. The main effect of time of testing was statistically significant (Wilks's $\Lambda = .77$), $F(2, 124) = 18.65$, $p < .0005$; pretest: $M = 16.01$, $SD = 5.69$; Posttest 1: $M = 18.79$, $SD = 5.44$; Posttest 2: $M = 18.68$, $SD = 5.94$. Students knew more about using comparative signaling words after instruction with ITSS than before completing ITSS, pretest to Posttest 1: $t(130) = 5.63$, $p < .0005$, $d = 0.49$; pretest to delayed posttest: $t(130) = 5.59$, $p < .0005$, $d = 0.49$. There was maintenance of performance on the signaling test 1 month after instruction with ITSS, $t(130) = 0.31$, $p = .757$. The main effect of reading ability also was statistically significant, $F(1, 123) = 26.41$, $p < .0005$; low readers: $M = 14.54$, $SD = 4.87$; middle readers: $M = 17.88$, $SD = 3.71$; high readers: $M = 21.08$, $SD = 3.54$. Tukey's honestly significant difference multiple comparisons tests showed statistically significant differences among all of the three reading comprehension groups in the use of comparison signaling words ($p = .001$).

The predicted individualization condition by time effect was statistically significant (Wilks's $\Lambda = .95$), $F(2, 124) = 0.95$, $p = .042$; see Table 11 for cell means and follow-up analyses for reading ability groups within conditions across times for testing. Students in both conditions made statistically significant gains from pretest to immediate posttest, individualized: $t(64) = 6.15$, $p < .0005$; standard: $t(65) = 2.16$, $p = .035$, and from pretest to delayed posttest, individualized: $t(64) = 4.78$, $p < .0005$; standard: $t(65) = 2.73$, $p = .008$. Neither condition showed decline

Table 10
Organization of Recall (TLS) for Three Reading Ability Groups

		Comparison TLS		Problem-and-solution TLS	
Reading ability	*n*	*M (SD)*	95% CI	*M (SD)*	95% CI
Low	43	4.89 (1.85)	[4.32, 5.45]	3.63 (1.69)	[3.09, 4.13]
Middle	44	6.24 (1.83)	[5.69, 6.80]	4.58 (1.72)	[4.00, 5.05]
High	44	6.61 (1.83)	[6.05, 7.16]	5.65 (2.13)	[5.09, 6.20]

Note. TLS = top-level structure; CI = confidence interval.

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

Table 11
Means (Standard Deviations) and Effect Sizes Between Conditions on the Signaling Test Over Testing Time

Individualization condition	*M (SD)* Pretest	Posttest 1	Posttest 2	*d*[a] Posttest 1–Pretest[b]	Posttest 2–Pretest	Posttest 2–Posttest 1
High ability readers						
Standard ($n = 22$)	19.55 (5.70)	21.55 (3.36)	21.97 (4.19)	0.35	0.42*	0.13
Individualized ($n = 22$)	17.64 (5.73)	22.73 (3.22)	22.91 (3.28)	0.89***	0.92***	0.06
Middle ability readers						
Standard ($n = 20$)	16.95 (4.65)	18.65 (4.57)	19.05 (5.87)	0.37	0.45	0.09
Individualized ($n = 24$)	15.42 (4.83)	19.00 (3.77)	18.21 (4.36)	0.74**	0.58*	−0.21
Low ability readers						
Standard ($n = 24$)	13.50 (6.14)	14.46 (6.00)	14.71 (6.17)	0.16	0.20	0.04
Individualized ($n = 19$)	12.95 (4.37)	16.42 (6.58)	15.21 (6.55)	0.79*	0.52	−0.18

[a] Difference between pretest and posttest, divided by the standard deviation on the pretest and difference between Posttest 2 and Posttest 1, divided by the standard deviation on Posttest 1. [b] Dependent *t* tests between two testing times.
* Statistically significant at $p < .05$. ** Significant at $p < .005$. *** Dependent *t* tests statistically significant at $p < .0005$.

from immediate to delayed posttest, individualized: $t(64) = 1.17$, $p = .245$; standard: $t(65) = 2.16$, $p = .511$. However, students in the individualized condition made more substantial gains on the signaling test from pretest to immediate posttest ($d = 0.78$) and pretest to delayed posttest ($d = 0.61$) than did students in standard ITSS from pretest to immediate posttest ($d = 0.25$) and pretest to delayed posttest ($d = 0.30$).

Did Individualization Affect Transfer to the Standardized Reading Comprehension Test?

A repeated measures analysis of variance was conducted on GSRT raw scores with individualization condition (individualized vs. standard), reading comprehension ability (high, middle, low), GSRT order (form A on pretest and B on posttest or the opposite order), and repeated measures on time of testing (before or after ITSS instruction) as predictor variables. There were two significant main effects and two significant interactions. Statistically significant main effects were time of testing (Wilks's $\Lambda = .78$), $F(1, 119) = 33.41$, $MSE = 63.04$, $p < .0005$, and reading ability (Wilks's $\Lambda = .78$), $F(1, 119) = 33.41$, $MSE = 138.99$, $p < .0005$. The hypothesized interaction between condition and time of testing was statistically significant (Wilks's $\Lambda = .97$), $F(1, 119) = 4.20$, $MSE = 63.04$, $p = .043$; see Table 12.[3] Also, there was a statistically significant interaction between reading ability and time of testing (Wilks's $\Lambda = .87$), $F(2, 119) = 8.71$, $MSE = 63.04$, $p < .0005$.

Statistically significant improvement in performance on the standardized test was demonstrated after structure strategy instruction (pretest: $M = 34.73$, $SD = 13.02$; posttest: $M = 40.20$, $SD = 11.90$). This improvement was significantly moderated by version of ITSS instruction. Table 12 displays the GSRT means, standard deviations, follow-up dependent *t* tests for the significant interaction, and effect sizes for the students who received the standard version of ITSS and those who received more individualized ITSS. Both ITSS groups made statistically significant gains between the pretest and posttest (see Table 12). The effect sizes shown in Table 12 can be helpful in interpreting the statistically significant interaction between individualized condition and time of testing. The students who received individualized ITSS made more substantial improvements from pretest to posttest on the GSRT ($d = 0.55$) than did students in the standard ITSS ($d = 0.30$). The effect size for the individualized group falls at the top of the medium range, whereas effects for the standard group can be classified as near the top of the small range (Cohen, 1988; Lipsey, 1990). Overall, the findings suggest that providing individualized order, number, and difficulty of lessons matched to the individual's performance in the ITSS lessons yields better outcomes than the standard order and difficulty of ITSS lessons.

The significant interaction between reading ability and time of testing is apparent from examining the means and effect sizes in Table 12. Lower ability readers made more improvements in reading comprehension than did higher ability readers. Large effect sizes between pretest and posttest performance on the GSRT were found for low ability ($d = 1.10$) and average ability readers ($d = 0.59$), but not for high ability readers ($d = 0.09$). The smaller pretest to posttest gains for students with high initial reading comprehension performance may be due to ceiling effects on the GSRT.

[3] All of the analyses for the dependent variables were rerun with covariates measuring number of texts read and number of 30-min sessions working online with ITSS. These covariates were examined to examine whether any differential gains with training from the two individualization conditions could be explained by time on task or extra practice reading texts. The covariates did not eliminate any of the hypothesized time by individualized condition interactions. The results from the analyses held regardless of the use of these covariates. For example in the analyses of the standardized reading comprehension test scores, the hypothesized interaction between condition and time of testing was still statistically significant (Wilks's $\Lambda = .96$), $F(1, 117) = 4.32$, $p = .040$; adjusted means for individualized group: pretest = 33.42 ($SD = 12.76$) and posttest = 41.51 ($SD = 12.76$); adjusted means for the standard group: pretest = 35.91 ($SD = 13.33$) and posttest = 39.35 ($SD = 11.87$).

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

Table 12
Interaction Between Standard Versus Individualized ITSS on Gray Silent Reading Test Pre- and Posttests

Individualization condition and reading ability	*M (SD)* Pretest	Posttest	Paired sample *t* test	d^a
Standard (n = 66)	35.41 (13.33)	39.35 (11.87)	$t(65) = 2.58, p = .012$	0.30
Low (n = 24)	23.58 (9.89)	31.62 (11.54)	$t(23) = 3.37, p = .003$	0.81
Middle (n = 20)	37.95 (9.47)	41.05 (9.38)	$t(19) = 1.16, p = .260$	0.33
High (n = 24)	46.00 (8.83)	46.23 (9.51)	$t(21) = .08, p = .934$	0.03
Individualized (n = 65)	34.05 (12.76)	41.06 (11.95)	$t(64) = 5.12, p < .0005$	0.55
Low (n = 19)	22.74 (8.51)	35.47 (11.10)	$t(18) = 6.39, p < .0005$	1.50
Middle (n = 24)	32.46 (8.87)	40.08 (11.68)	$t(23) = 3.42, p = .002$	0.86
High (n = 22)	45.55 (9.55)	46.95 (10.71)	$t(21) = .62, p = .541$	0.15

[a] Difference between pretest and posttest divided by the standard deviation on the pretest.

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

Did Greater Individualization Affect Mastery Goals and Work in the Lessons?

Mastery goals. The individualized ITSS was hypothesized to increase students' mastery–learning goals more than standard ITSS. Learning goals were expected to develop from instruction better matched to a student's skill level (individualized ITSS) rather than instruction that was too hard or easy (standard ITSS). The reduction in mismatches between skill and instruction levels was hypothesized to provide learning situations for each student that were more conducive for developing mastery goals while working in ITSS. No differences between conditions were expected for performance goals (approach or avoidance). A MANOVA was conducted on the three goal orientation scales from the PALS measure. Variables examined were individualization condition (individualized vs. standard ITSS) and reading comprehension ability (high, middle, and low). The only statistically significant finding was the main effect for individualization condition (Wilks's $\Lambda = .91$), $F(3, 123) = 4.06$, $p = .009$. Reading ability was not statistically significant (Wilks's $\Lambda = .95$), $F(6, 246) = 1.06$, $p = .388$, nor was the interaction between reading ability and individualization condition (Wilks's $\Lambda = .96$), $F(6, 246) = 0.81$, $p = .565$. Cell means and standard deviations are displayed in Table 13.

As predicted, there was a statistically significant effect for individualization condition on mastery goals, $F(1, 131) = 10.92$, $MSE = 23.33$, $p = .001$. The average mastery goals score for the individualized instruction group was 19.39 ($SD = 4.39$), whereas the average for the group receiving standard ITSS was 16.60 ($SD = 5.28$; $d = 0.53$). As predicted, condition did not affect performance goals; approach: $F(1, 131) = 0.002$, $MSE = 25.52$, $p = .965$; avoidance: $F(1, 131) = 0.14$, $MSE = 17.95$, $p = .708$.

There was no pretest on mastery goals because we thought it was unreasonable to ask students about learning goals during ITSS prior to starting the ITSS lessons. Because students were randomly assigned to the individually tailored ITSS and standard ITSS, it appears that the ITSS with more individualized instruction caused students to be more engaged in the lessons. To further investigate this causal claim, we ran the analysis again with pretest scores on the GSRT as a covariate. Again, type of instruction was statistically significant, $F(1, 128) = 11.65$, $MSE = 23.22$, $p = .001$. The adjusted means were 19.44 for the more individually tailored group and 16.56 for standard ITSS group. To further examine the relation between achievement motivation to master ITSS content and individualization condition, we examined performance in the lessons.

Performance in the lessons. An analysis of covariance (ANCOVA) was conducted, comparing individualization condition (individualized or standard ITSS) on the number of student responses in the comparison lessons achieving the criteria of 50% or better and earning a feedback response of "good" from I.T. (see Table 3 regarding scoring criteria in ITSS). The 12 comparison lessons were selected for examination. The total number of responses submitted by the students in the comparison lessons was used as a covariate because students in the individualized group had the potential of receiving more comparison lessons for remediation or enrichment. There was a statistically significant effect for type of ITSS instruction, $F(1, 128) = 10.36$, $MSE = 48.50$, $p = .002$. Students using individually tailored ITSS submitted more good responses in the lessons ($M = 31.43$, $SD = 6.99$) than did

Table 13
Means (Standard Deviations), and Effect Sizes for Achievement Goals on the Posttest for Individualization Conditions and Reading Ability Levels

Mastery goals and reading ability	Individualization conditions *M (SD)* Standard (n = 66)	Individualized (n = 65)	d^a
Work mastery	16.60 (5.28)	19.39 (4.39)	0.53
Low (n = 43)	16.88 (4.87)	20.24 (3.64)	0.69
Middle (n = 44)	18.10 (5.00)	19.16 (4.91)	0.21
High (n = 44)	14.94 (5.71)	18.91 (4.47)	0.70
Performance approach	14.66 (4.99)	14.68 (5.03)	0.004
Low (n = 43)	14.41 (5.04)	14.37 (4.60)	−0.01
Middle (n = 44)	15.86 (4.83)	14.83 (5.26)	−0.21
High (n = 44)	13.83 (5.10)	14.78 (5.35)	0.19
Performance avoidance	12.62 (4.27)	12.37 (4.14)	−0.06
Low (n = 43)	12.71 (4.01)	12.33 (4.45)	−0.09
Middle (n = 44)	12.92 (3.70)	13.19 (4.15)	0.07
High (n = 44)	12.24 (5.11)	11.51 (3.86)	−0.14

[a] Difference between standard and individualized Intelligent Tutoring of the Structure Strategy divided by the standard deviation for standard.

students in standard ITSS (M = 27.65, SD = 6.98). The average adjusted score for the group receiving more individualized instruction was 31.51, whereas it was 27.58 for the group receiving standard ITSS.

Did Greater Individualization Affect Attitudes Toward Computers or Self-Efficacy?

Reading comprehension level did not interact with differentiation condition on any of the reading comprehension or achievement goal measures examined previously in this study. To simplify presentation of the remainder of the findings, only the factors of differentiation condition and time of testing will be examined for the computer attitudes and self-efficacy measures.

Computer attitudes. For the computer attitudes questionnaire data, a repeated-measures ANOVA was conducted with repeated measures on the time of testing (pretest vs. immediate posttest) and the predictor variable of individualization condition (individualized vs. standard ITSS). There was no main effect for condition, $F(1, 129) = 0.77$, $MSE = 201.12$, $p = .381$, but there was a statistically significant effect for time of testing (Wilks's $\Lambda = .93$), $F(1, 129) = 9.60$, $p = .002$. As seen in Table 14, attitudes toward computers decreased from the pretest to the posttest. Although there is no control group for comparison, the decline is probably caused by ITSS. The data match our observation that prior to ITSS instruction, students generally associated the computer with games and fun. In contrast, ITSS instruction involved working hard on the computer to learn new reading skills.

The data suggest that the more individualized version of ITSS showed less of this decline in attitudes toward computers. There was a trend for steeper decline in attitudes with standard ITSS than for the individualized ITSS, $F(1, 129) = 3.66$, $MSE = 56.82$, $p = .058$. Observed power for the Time × Condition interaction was low (.47). As seen in Table 14, the average posttest score for standard ITSS declined about 5 points, whereas the decline was only about 1 point for individualized ITSS. A simple ANCOVA with computer attitudes pretest scores as the covariate examined effects for condition on posttest computer attitudes. For this analysis, total pretest and posttest scores were divided by the number of items on the questionnaire (21) in order to aid in interpretation of the results on the 5-point Likert scale. Findings from the ANCOVA showed higher adjusted scores for individualized ITSS (M = 3.67, SD = 0.54) than for standard ITSS (M = 3.51, SD = 0.62); $F(1, 128) = 3.90$, $MSE = 0.23$, $p = .050$. The average Likert scale scores on the pretest was 3.72 (SD = 0.52) for individualized ITSS and 3.73 (SD = 0.47) for standard ITSS. For both groups, the average pretest scores fell closer to *agree somewhat* (4) with positive statements about computers than *indifference* (3). For the individualized group, there was no decline from pretest to posttest, $t(64) = 0.90$, $p = .372$. There was significant decline and movement toward indifference for standard ITSS, $t(65) = 3.34$, $p = .001$. The results suggest that changes to ITSS for greater individualization can improve reading comprehension skills while maintaining a tendency toward somewhat positive attitudes toward computers.

Self-efficacy. To examine self-efficacy, a repeated-measures ANOVA was conducted with repeated measures on time of testing (pretest self-efficacy scores vs. immediate posttest scores) and the predictor variable of individualization condition (individualized vs. standard ITSS). Contrary to the findings of Meyer et al. (2002) with human tutors, there were no gains in self-efficacy from pretest to posttest (Wilks's $\Lambda = .995$), $F(1, 129) = 0.70$, $p = .404$, and no significant interaction between time and condition (Wilks's $\Lambda = 1.00$), $F(1, 129) = 0.012$, $p = .915$; see Table 14. Providing students with more individualized lessons did not promote greater general self-efficacy. In the Meyer et al. (2002) study, the self-efficacy of both students and tutors increased. The current findings suggest that these increases may have resulted from successful human interactions rather than increased reading comprehension skills.

Table 14
Means and Standard Deviations and Effect Sizes for Individualization Conditions on Computer Attitudes and Self-Efficacy at Pretest and Posttest

Measure and individualization condition	M (SD) Pretest	Posttest	d[a]
Computer attitudes	78.20 (10.41)	75.30 (12.26)	−0.28
Standard (n = 66)	78.32 (9.90)	73.65 (12.93)	−0.47
Individualized (n = 65)	78.08 (10.99)	76.97 (11.40)	−0.10
Self-efficacy	72.54 (9.77)	73.27 (9.57)	0.07
Standard (n = 66)	71.85 (10.47)	72.68 (10.09)	0.08
Individualized (n = 65)	73.24 (9.02)	73.88 (9.05)	0.07

[a] Difference between pretest and posttest divided by the standard deviation on the pretest.

Discussion

In this study, we examined the effects of different versions of a web-based tutoring system to teach the structure strategy. The design feature varied was individualization of instruction, and it was assigned through stratified random assignment to fifth-grade students. The primary research question was whether the two different versions (individualized ITSS and standard ITSS) affected reading comprehension. The individualized version was designed to match the individual learner's needs by using the student's performance during a lesson to adapt the sequence, complexity, and/or text difficulty of the proceeding lesson. The standard ITSS version provided students with the same, fixed sequence of lessons regardless of their performance. Reading comprehension was assessed by students recalling information and filling in signaling words for researcher-designed texts as near transfer measures and by the GSRT, a standardized test, as the far transfer measure. Results showed that individualized ITSS substantially increased reading comprehension on the standardized test over standard ITSS. The predicted interaction between individualization condition and time of testing was found on the GSRT and the signaling test, but not the free recall measures.

For the free recall task, students in both individualized and standard ITSS conditions made similar gains from pretest to the two posttests in the amount of information they could remember from the text and the organization of their recalls in terms of the TLS (e.g., pretest to immediate posttest: d = 0.75 to 1.45; pretest to 1-month delayed posttest: d = 0.63 to 1.43). The recall measures tested aspects of the strategy that were explicitly taught and practiced in most lessons. This explicit instruction and repetition

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

may have been sufficient to dramatically boost performance on these measures regardless of whether the lessons were individualized. The large gains for both versions of ITSS and the large variance typically associated with free recall data may have meant that our procedures were not sensitive enough to detect the differences that were found on the GSRT and the signaling test.

The interaction between condition (individualized vs. standard) and time of testing (pretest vs. immediate posttest vs. delayed posttest) was statistically significant for the signaling test. Students in the individualized condition made more substantial gains in use of comparative signaling words from pretest to immediate posttest ($d = 0.78$) and pretest to delayed posttest ($d = 0.61$) than students in standard ITSS from pretest to immediate posttest ($d = 0.25$) and pretest to delayed posttest ($d = 0.30$). Unlike the free recall, the signaling test involved a novel format for testing knowledge of signaling words that had not been used in the ITSS instruction. Both versions of ITSS explicitly taught signaling and evaluated knowledge of signaling words within the lessons by (a) clicking on signaling words in texts and (b) using signaling primarily when composing main idea statements or recalls, but not by filling in blanks for signaling words in texts. The signaling test was also a generative task in that students had to understand the concept of signaling and then generate their own comparison signaling words to fill in blanks in the text (see Appendix B).

Meyer et al. (2010) found that the design feature of feedback did not impact the signaling test. The feedback (elaborated vs. simple) conditions varied in whether students were reminded to consult their signaling tables and add signaling words to main idea statements or recalls. However, in the current study, the targeted design feature did impact performance on the signaling test, and individualized instruction may have resulted in a more thorough learning of signaling words. Students in the individualized condition, who received both the standard and alternate topic lessons (e.g., remediation: lesson 15 and 15a), were provided with practice in identifying the same signaling words in different versions of the texts. This repetition of signaling with different content may have promoted increasing understanding and transfer of learning about signaling. For students receiving enrichment, signaling words cuing the organization of less familiar text (e.g., chinchillas vs. pot-bellied pigs) may have yielded more learning about signaling than acquired by similar students in the standard condition who read about familiar context (e.g., dogs vs. cats) and could rely less on signaling for understanding.

There was clear support for the hypothesized greater performance in reading comprehension on the standardized test for students who completed individualized ITSS. On the standardized GSRT measure, students who received individualized ITSS showed more improvement ($d = 0.55$) than students who received standard ITSS ($d = 0.30$). This result indicates that individually tailored structure strategy instruction with expository text transfers to substantial improvements on a standardized reading comprehension test. Effects of interventions on standardized reading comprehension tests are not commonplace (e.g., Gamse, Bloom, Kemple, & Jacobs, 2008). This finding also supports and extends the work of Kalyuga and Sweller (2005), which compared individualized and fixed versions of an algebra tutor for tenth-grade students and reported superior algebraic gains for individualized instruction. The current investigation answered the challenge of Kalyuga and Sweller for further research with less structured domains, particularly individualizing instruction based on online comprehension performance with reading tutors. The rapid online assessment and resultant individualization of lessons in both studies improved skills with different domains and ages of learners.

Three secondary questions were also investigated in this study. The first question was whether the variation of instruction affected learning goals while working in ITSS, quality of work in lessons, or attitudes toward computers or self. The results indicated that the individually tailored ITSS group showed higher learning goals while working in ITSS lessons and better quality of work in the lessons than the standard instruction group. This finding of better work in lessons supports previous findings of a study of individualized training in air traffic control (Camp et al., 2001), in which students receiving training based on their performance did better on task related measures collected during training than students receiving fixed training.

After adjusting posttest scores on the basis of initial attitudes toward computers, students in individualized ITSS had more positive attitudes toward computers than students in standard ITSS. There was a trend for steeper pretest-to-posttest declines in attitudes toward computers for the standard ITSS group than for the individually tailored group. Changes in ITSS that can maintain positive attitudes toward computers while increasing reading comprehension are important and worth pursuing in future research.

Instruction matched to a student's skill level resulted in higher learning goals, work quality, and attitudes toward computers, but there were no significant differences between the groups on general self-efficacy. Meyer et al. (2002) found that after completing web-based tutoring, the self-efficacy of both fifth-grade students and their older adult tutors increased. The current findings suggest that these increases may have resulted from successful human online interactions rather than increased reading comprehension skills. Future research is needed to further investigate self-efficacy with more domain specific measures of reading self-efficacy (e.g., Schunk & Rice, 1987) rather than the general measure used in this study.

The next secondary research question focused on gains from pretest to posttest on the experimenter-made materials and whether they were maintained 1 month after instruction. Large gains on the near transfer measures were made for both individualization conditions. Findings indicated that gains made immediately after ITSS were maintained 1 month after instruction for most measures. Similarly, Meyer et al. (2010) generally found maintenance over summer break after ITSS instruction.

The final secondary question was whether our attempt to better match the instruction students received to their online performance similarly impacted students that were initially high, middle, or low on the standardized reading comprehension test. All three reading ability groups benefited more from individualized ITSS in comparison with standard ITSS. Substantial gains from pretest to posttest on the GSRT were found for low ability readers ($d = 1.10$) and middle ability readers ($d = 0.59$), but not for high ability readers ($d = 0.09$). The smaller pretest to posttest gains for students with high initial reading comprehension performance may be due to ceiling effects on the GSRT. Unlike the GSRT, statistically significant interactions between reading comprehension ability by time of testing were not found on the recall or signaling measures in which ceiling effects were not an issue (e.g., $d = 0.89$ between pretest and posttest for high ability readers on the signal-

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

ing test). The results indicated that the superiority of individualized ITSS over standard ITSS held similarly across ability levels.

Findings from the current investigation support the theory that more learner appropriate scaffolding leads to better learning (e.g., O'Donnell et al., 2007; Vygotsky, 1978). The interaction between condition and time of testing on the signaling test and far transfer task (GSRT) supports the explanation that the gains in reading performance are primarily due to better learning in individualized ITSS, which provided lessons that were neither too easy nor too difficult for students. The data about students' learning goals while working in ITSS and the quality of their work in the lessons also suggest that better learning resulted from individualized ITSS than standard ITSS. Instruction closer to performance levels also appeared to better maintain somewhat positive attitudes toward the mode of the instruction, the computer.

We hypothesized that the design feature of individualization would have both a cognitive and a motivational effect. A more complicated model was not predicted, such as individualization affecting learning achievement goals that in turn affect cognitive skills. Post hoc investigations with ANCOVA did not support a more complicated model. For example, when mastery goals were used as a covariate in a repeated-measures ANCOVA on pretest and posttest GSRT scores, the hypothesized interaction between time of testing and individualization condition held (Wilks's Λ = .96), $F(1, 112) = 5.14$, $MSE = 63.45$, $p = .025$.

A post hoc look at three experimenter-designed yes-or-no questions helps to capture the students' perspective on the study. Three questions were posed via the computer at two times: first, after students had completed pretesting and were first introduced to ITSS and again, during the last computer session of ITSS. The questions (whether or not a student agreed with a statement) were (a) "I enjoyed participating in this study"; (b) "I understood the deeper meanings of the text"; and (c) "I was highly motivated in this study." There were no significant differences between the individualized conditions at pretest for each of the three questions, $\chi^2(1, N = 123) = 0.02$, $p = .900$; $\chi^2(1, N = 101) = 1.65$, $p = .199$; and χ^2 $(1, N = 123) = 1.70$, $p = .192$, respectively, but there were significant differences on each question at posttest. For the question about enjoying the study, 81% of the students in the individualized condition responded yes, whereas 61% responded yes in the standard condition, $\chi^2(1, N = 123) = 5.94$, $p = .015$). For the question about understanding the deeper meanings of the text, 84% of the students in the individualized condition said yes, whereas 67% said yes in the standard condition, $\chi^2(1, N = 123) = 4.6$, $p = .031$). For the final question about motivation, 60% of the students in the individualized condition indicated yes, whereas 38% indicated yes in the standard condition for the three questions, $\chi^2(1, N = 123) = 5.94$, $p = .015$. These responses mirror the findings of positive effects of individualization on attitudes, reading comprehension, and motivation. Overall, the smaller jumps in complexity for practice lessons, the greater flexibility in the sequence of lessons, and the reduced difficulty of texts appear to have resulted in greater success in the lessons, better reading comprehension, mastery achievement goals, and more positive attitudes.

The task selection approach in the more individualized ITSS can be categorized as a dynamic whole-task selection approach (Salden, Paas, & van Merrienboer, 2006). Salden et al. (2006) compared different task selection approaches in the training of complex cognitive skills (i.e., dynamic versus static and part versus whole). Dynamic approaches can be adjusted to the needs of the individual during training (i.e., individualized ITSS), whereas static approaches cannot be so adjusted because the order and complexity of the training tasks are predetermined prior to training (i.e., standard ITSS). Whole approaches are usually needed for complex cognitive tasks, such as the structure strategy, in which parts are related and less amenable to teaching in isolation before being combined. Dynamic whole-task selection approaches present complete learning tasks and adjust the order and complexity of the learning tasks based on the needs of the individual student during training. These characteristics make this approach highly flexible and adaptive, optimizing the training processes. The current study's findings of the significant gains on an unpracticed format (the signaling test) and the far transfer GSRT support the advantages of the dynamic whole-task selection approaches.

The study showed that students who received the more individualized instruction made greater gains on the far transfer measure of reading comprehension and the signaling test than did those in standard instruction. However, on most of the near transfer measures that were similar and frequently used in both versions of ITSS, there was no significant difference between the two conditions. Further investigations are needed to look into which aspects of the more individualized instruction contributed to these substantial gains beyond those of standard ITSS. Although it was not possible in this study to ascertain which specific components of individualized ITSS contributed to students' greater success on the GSRT, the findings have implications for the small but growing body of research about digital learning environments and intelligent tutoring systems designed to improve students' reading comprehension (e.g., McNamara et al., 2007). Future research will need to specify which ITSS lessons and steps for individualizing are the most beneficial. The current investigation shows that nonhuman, pedagogical agents can improve reading comprehension and that they can do so more effectively if they individualize instruction.

Generalization of the study's findings to other settings requires further verification due to limitations of the study. External validity limitations include a volunteer sample from only one school district and attrition from this sample. Also, the study was conducted with students and teachers familiar with computers, so the findings may not generalize to other school settings. In terms of internal validity, the design was robust regarding the individualization variable. However, limitations are inherent in the study's pretest–posttest design without a control group (Campbell & Stanley, 1963). These limitations result in other possible explanations of the substantial pretest to posttest gains in reading comprehension, such as maturation.

However, our first web-based intervention teaching the structure strategy (Meyer et al., 2002) did have a control group, and similar pretest to posttest gains were reported for the structure strategy intervention as that found in the current study. For example with regard to TLS scores, the percentage of children in the control using the same TLS as used in the text for at least one of the two structures (comparison and problem & solution) during a pretest, immediate posttest, and 2 1/2-month delayed posttest was 40%, 50%, and 35%, respectively. In contrast, percentages for use of the strategy for the students receiving online structure strategy instruction were 35%, 75%, and 85%, respectively. In the current study, percentages for students were 41%, 80%, and 77% for the pretest, immediate posttest, and 1-month delayed posttest, respectively.

Some could argue that the large reading comprehension gains are due in part to I.T. orally reading most of the texts to students when the texts were first introduced. Repeated oral reading has been particularly helpful in developing fluency for beginning readers, but not as consistently for older struggling readers (Wexler, Vaughn, Roberts, & Denton, 2010). It is important to note that similar gains in structure strategy use and recall were found in the Meyer et al. (2002) study with no oral component. This similarity argues against the oral reading explanation for increases in reading comprehension. Additionally, both conditions had the same oral reading component, but substantially larger gains were found on the standardized test for the individualized condition. The oral component was introduced to pace students through a lesson, prevent skipping of important instructions, and enable poor readers to learn the strategy through oral language skills rather than limiting the students to what they could decode from the screen.

This study is important because it highlights the importance of individualizing lesson sequence, difficulty, and complexity by using students' within lesson performance. Individualizing instruction in intelligent tutoring systems, like ITSS, is helpful with particular reading tasks, such as those found on the standardized reading comprehension test used in this study. Further individualization of ITSS instruction may yield promising theoretical and practical results. Future research needs to increase our understanding of how much and what types of individualized instruction are most effective across a variety of instructional tasks, domains, and readers.

References

Aleven, V., Ashley, K., Lynch, C., & Pinkwart, N. (2008). Intelligent tutoring systems for ill-defined domains: Assessment and feedback in ill-defined domains. *Proceedings of the Workshop on Intelligent Tutoring Systems for Ill-Defined Domains at the 9th International Conference on Intelligent Tutoring Systems.* Montréal, Quebec, Canada. Retrieved from http://www.cs.pitt.edu/~collinl/ITS08/

Ames, C. (1992). Classrooms: Goals, structures, and student motivation. *Journal of Educational Psychology, 84,* 261–271. doi:10.1037/0022-0663.84.3.261

Ames, C., & Archer, J. (1988). Achievement goals in the classroom: Students' learning strategies and motivation processes. *Journal of Educational Psychology, 80,* 260–267. doi:10.1037/0022-0663.80.3.260

Anderman, L. H. (1999). Classroom goal orientation, school belonging and social goals as predictors of students' positive and negative affect following the transition to middle school. *Journal of Research and Development in Education, 32,* 89–103.

Armbruster, B. B., Anderson, T. H., & Ostertag, J. (1987). Does text structure/summarization instruction facilitate learning from expository text? *Reading Research Quarterly, 22,* 331–346. doi:10.2307/747972

Baddeley, A. (1992). Working memory: The interface between memory and cognition. *Journal of Cognitive Neuroscience, 4,* 281–288. doi:10.1162/jocn.1992.4.3.281

Bartlett, B. J. (1978). *Top-level structure as an organizational strategy for recall of classroom text* (Unpublished doctoral dissertation). Arizona State University, Tempe.

Blok, H., Oostdam, R., Otter, M. E., & Overmaat, M. (2002). Computer-assisted instruction in support of beginning reading instruction: A review. *Review of Educational Research, 72,* 101–130. doi:10.3102/00346543072001101

Britton, B. K., & Graesser, A. C. (Eds.). (1996). *Models of understanding text.* Mahwah, NJ: Erlbaum.

Camp, G., Paas, F., Rikers, R., & van Merrienboer, J. (2001). Dynamic problem selection in air traffic control training: A comparison between performance, mental effort and mental efficiency. *Computers in Human Behavior, 17,* 575–595. doi:10.1016/S0747-5632(01)00028-0

Campbell, D. T., & Stanley, J. C. (1963). Experimental and quasi-experimental designs for research on teaching. In N. L. Gage (Ed.), *Handbook of research on teaching* (pp. 171–246). Chicago, IL: Rand McNally.

Chall, J. S., & Dale, E. (1995). *Readability revisited: The new Dale-Chall readability formula.* Cambridge, MA: Brookline Books.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). New York, NY: Academic Press.

Cook, L. K., & Mayer, R. E. (1988). Teaching readers about the structure of scientific text. *Journal of Educational Psychology, 80,* 448–456.

Corno, L. (2008). On teaching adaptively. *Educational Psychologist, 43,* 161–173. doi:10.1080/00461520802178466

Corno, L., Cronbach, L. J., Kupermintz, H., Lohman, D. F., Mandinach, E. B., Porteus, A. W., & Talbert, J. E. (2002). *Remaking the concept of aptitude: Extending the legacy of Richard E. Snow.* Mahwah, NJ: Erlbaum.

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

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

Dale, E., & Chall, J. S. (1948). A formula for predicting readability. *Educational Research Bulletin, 27,* 11–28.

Dalton, B., & Strangman, N. (2006). Improving struggling readers' comprehension through scaffolded hypertexts and other computer-based literacy programs. In M. C. McKenna, L. D. Labbo, R. D. Kieffer, & D. Reinking (Eds.), *International handbook of literacy and technology* (Vol. 2, pp. 75–92). Mahwah, NJ: Erlbaum.

Day, E. A., Radosevich, D. J., & Chasteen, C. S. (2003). Construct- and criterion-related validity of four commonly used goal orientation instruments. *Contemporary Educational Psychology, 28,* 434–464. doi:10.1016/S0361-476X(02)00043-7

Dweck, C. S. (1986). Motivational processes affecting learning. *American Psychologist, 41,* 1040–1048. doi:10.1037/0003-066X.41.10.1040

Elliot, A. J., & Church, M. A. (1997). A hierarchical model of approach and avoidance achievement motivation. *Journal of Personality and Social Psychology, 72,* 218–232. doi:10.1037/0022-3514.72.1.218

Englert, C. S., & Hiebert, E. H. (1984). Children's developing awareness of text structures in expository materials. *Journal of Educational Psychology, 76,* 65–74. doi:10.1037/0022-0663.76.1.65

Flesch, R. (1948). A new readability yardstick. *Journal of Applied Psychology, 32,* 221–233. doi:10.1037/h0057532

Gambrell, L. B., Morrow, L. M., & Pressley, M. (2007). *Best practices in literacy instruction* (3rd ed.). New York, NY: Guilford Press.

Gamse, B. C., Bloom, H. S., Kemple, J. J., & Jacobs, R. T. (2008). *Reading first impact study: Interim report* (NCEE 2008–4016). Washington, DC: National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education. Retrieved from http://ies.ed.gov/ncee/pdf/20084016.pdf

Gernsbacher, M. A., Varner, K. R., & Faust, M. K. (1990). Investigating differences in general comprehension skill. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 16,* 430–445. doi:10.1037/0278-7393.16.3.430

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

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

puters, 36, 180–192. Retrieved from http://brm.psychonomic-journals.org/

Graesser, A. C., Singer, M., & Trabasso, T. (1994). Constructing inferences during narrative text comprehension. *Psychological Review, 101,* 371–395. doi:10.1037/0033-295X.101.3.371

Gregory, G. H., & Chapman, C. (2002). *Differentiating instructional strategies: One size doesn't fit all.* Thousand Oaks, CA: Corwin Press.

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

Harackiewicz, J. M., Barron, K. E., Pintrich, P. R., Elliot, A. J., & Thrash, T. M. (2002). Revision of achievement goal theory: Necessary and illuminating. *Journal of Educational Psychology, 94,* 638–645. doi:10.1037/0022-0663.94.3.638

Kalyuga, S., & Sweller, J. (2005). Rapid dynamic assessment of expertise to improve the efficiency of adaptive e-learning. *Educational Technology, Research and Development, 53,* 83–93. doi:10.1007/BF02504800

Kintsch, W. (1998). *Comprehension: A paradigm for cognition.* New York, NY: Cambridge University Press.

Kintsch, W., & van Dijk, T. A. (1978). Toward a model of text comprehension and production. *Psychological Review, 85,* 363–394. doi:10.1037/0033-295X.85.5.363

Kolen, M. J., & Brennan, R. L. (1995). *Test equating: Methods and practices.* New York, NY: Springer-Verlag.

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

Leong, C. K. (1995). Effects of on-line reading and simultaneous DECtalk auding in helping below-average and poor readers comprehend and summarize text. *Learning Disability Quarterly, 18,* 101–116. doi:10.2307/1511198

Lexile. (2005). The Lexile framework for reading: Lexile analyzer [Computer software]. Retrieved from www.lexile.com/DesktopDefault.aspx?view=re&tabindex=2&tabid=31&tabpageid=358

Lipsey, M. W. (1990). *Design sensitivity: Statistical power for experimental research.* Newbury Park, CA: Sage.

Mandler, J. M., & Johnson, N. S. (1977). Remembrance of things parsed: Story structure and recall. *Cognitive Psychology, 9,* 111–151. doi:10.1016/0010-0285(77)90006-8

Mann, W. C., & Thompson, S. A. (1986). Relational propositions in discourse. *Discourse Processes, 9,* 57–90. doi:10.1080/01638538609544632

McNamara, D. S., Louwerse, M. M., Cai, Z., & Graesser, A. (2005). Coh-Metrix (Version 1.4) [Computer software]. Retrieved from http://cohmetrix.memphis.edu

McNamara, D. S., O'Reilly, T., Rowe, M., Boonthum, C., & Levistein, I. (2007). iSTART: A web-based tutor that teaches self-explanation and metacognitive reading strategies. In D. S. McNamara (Ed.), *Reading comprehension strategies: Theories, interventions, and technologies* (pp. 397–420). Mahwah, NJ: Erlbaum.

Meece, J. L., Blumenfeld, P. C., & Hoyle, R. H. (1988). Students' goal orientations and cognitive engagement in classroom activities. *Journal of Educational Psychology, 80,* 514–523. doi:10.1037/0022-0663.80.4.514

Meyer, B. J. F. (1975). *The organization of prose and its effects on memory.* Amsterdam, the Netherlands: 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. doi:10.1097/00011363-200307000-00007

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. doi:10.2307/747349

Meyer, B. J. F., & McConkie, G. W. (1973). What is recalled after hearing a passage? *Journal of Educational Psychology, 65,* 109–117. doi:10.1037/h0034762

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. doi:10.1037/0022-0663.94.3.486

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. doi:10.1037/0022-0663.93.1.141

Meyer, B. J. F., & Rice, G. E. (1984). The structure of text. In P. D. Pearson, R. Barr, M. L. Kamil, & P. Mosenthal (Eds.), *Handbook of reading research* (pp. 319–351). New York, NY: Longman.

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

Meyer, B. J. F., & Wijekumar, K. (2007). A web-based tutoring system for the structure strategy: Theoretical background, design, and findings. In D. S. McNamara (Ed.), *Reading comprehension strategies: Theories, interventions, and technologies* (pp. 347–374). Mahwah, NJ: Erlbaum.

Meyer, B. J. F., Wijekumar, K., Middlemiss, W., Higley, K., Lei, P., Meier, C., & Spielvogel, J. (2010). Web-based tutoring of the structure strategy with or without elaborated feedback or choice for fifth- and seventh-grade readers. *Reading Research Quarterly, 45,* 62–92. doi:10.1598/RRQ.45.1.4

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: Erlbaum.

Midgley, C., Maehr, M. L., Hruda, L. Z., Anderman, E., Anderman, L., Freeman, K. E., & Urdan, T. (2000). *Manual for the Patterns of Adaptive Learning Scales (PALS*). Ann Arbor, MI: University of Michigan. Retrieved from http://www.umich.edu/~pals/PALS%202000_V12Word97.pdf

Miller, G. A. (2005). *Wordnet* [Database]. Retrieved from http://wordnet.princeton.edu/

Murphy, P. K., & Alexander, P. A. (2000). A motivated exploration of motivation terminology. *Contemporary Educational Psychology, 25,* 3–53. doi:10.1006/ceps.1999.1019

O'Donnell, A. M., Reeve, J., & Smith, J. K. (2007). *Educational psychology: Reflection for action.* Hoboken, NJ: Wiley.

Peng, C.-Y. J., Harwell, M., Liou, S.-M., & Ehman, L. H. (2007). Advances in missing data methods and implications for educational research. In S. S. Sawilowsky (Ed.), *Real data analysis* (pp. 31–78). Charlotte, NC: Information Age.

Penn Treebank. (2005). [Database]. Retrieved from www.cis.upenn.edu/~treebank/home.html

Pintrich, P. R. (2000). An achievement goal theory perspective on issues in motivation terminology, theory, and research. *Contemporary Educational Psychology, 25,* 92–104. doi:10.1006/ceps.1999.1017

Raphael, T. E., & Kirschner, B. M. (1985). The effects of instruction in compare/contrast text structure on sixth-grade students' reading comprehension and writing products (*Research Series No. 161).* East Lansing, MI: Michigan State University, Institute for Research on Teaching.

Reinking, D. (1988). Computer-mediated text and comprehension differences: The role of reading time, reader preference, and estimation of learning. *Reading Research Quarterly, 23,* 484–498. Retrieved from http://www.reading.org/General/Publications/Journals/RRQ.aspx doi:10.2307/747645

Reinking, D., & Schreiner, R. (1985). The effects of mediated text on measures of reading comprehension and reading behavior. *Reading Research Quarterly, 20,* 536–552. doi:10.2307/747941

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

Reutzel, D. R. (2007). Organizing effective literacy instruction: Differentiating instruction to meet the needs of all children. In L. B. Gambrell, L. M. Morrow, & M. Pressley (Eds.), *Best practices in literacy instruction* (3rd ed., pp. 313–343). New York, NY: Guilford Press.

Reynolds, C. R., & Kamphaus, R. W. (Eds.). (2003). *Handbook of psychological and educational assessment of children* (2nd ed.). New York, NY: Guilford.

Salden, R. J. C. M., Paas, F., & van Merrienboer, J. J. G. (2006). A comparison of approaches to learning task selection in the training of complex cognitive skills. *Computers in Human Behavior, 22,* 321–333. doi:10.1016/j.chb.2004.06.003

Sanders, T. J. M., Spooren, W. P., & Noordman, L. G. M. (1992). Toward a taxonomy of coherence relations. *Discourse Processes, 15,* 1–35. doi:10.1080/01638539209544800

Schulz, E. M., & Nicewander, W. A. (1997). Grade equivalent and IRT representations of growth. *Journal of Educational Measurement, 34,* 315–331. doi:10.1111/j.1745-3984.1997.tb00521.x

Schunk, D. H., & Rice, J. M. (1987). Enhancing comprehension skill and self-efficacy with strategy value information. *Journal of Reading Behavior, 19,* 285–302. doi:10.1080/10862968709547605

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

SPSS. (2007). *SPSS missing value analysis 16.0.* Chicago, IL: SPSS.

Test Your Document's Readability. (2007). In Microsoft Office Online. Retrieved from http://office.microsoft.com/en-us/help/HP101485061033.aspx

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

van den Broek, P., Young, M., Tzeng, Y., & Linderholm, T. (1999). The landscape model of reading: Inferences and the online construction of a memory representation. In S. R. Goldman & H. van Oostendorp (Eds.), *The construction of mental representations during reading* (pp. 71–98). Mahwah, NJ: Erlbaum.

Vygotsky, L. S. (1978). *Mind in society: The development of higher psychological processes.* Cambridge, MA: Harvard University Press.

Wentzel, K. R. (1999). Social-motivational processes and interpersonal relationships: Implications for understanding motivation at school. *Journal of Educational Psychology, 91,* 76–97. doi:10.1037/0022-0663.91.1.76

Wexler, J., Vaughn, S., Roberts, G., & Denton, C. A. (2010). The efficacy of repeated reading and wide reading practice for high school students with severe reading disabilities. *Learning Disabilities Research & Practice, 25,* 2–10. doi:10.1111/j.1540-5826.2009.00296.x

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

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. doi:10.1037/a0013152

Woolf, B. P. (2009). *Building intelligent interactive tutors: Student-centered strategies for revolutionizing e-learning.* Retrieved from http://books.google.com/books

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

Appendix A

Table A1

Lesson Order, Content, and Activities for Standard ITSS and Text Variations Available for Lessons in Individualized ITSS

		Available texts for individualized ITSS	
Standard order	Lesson content in standard ITSS	Alternative text topic for lesson in adjacent left column	Easier versions[a]
	Understanding the structure strategy with focus on comparison structure		
1 & 2	I.T. models structure strategy: comparison		
3	Compare two types of elephants: SI, *s*M, *sR*	Compare two types of crocodilians	Yes (2)
4	Easy text comparing two whales: SI, *s*M, *s*R	Easy text comparing two bears types	
5	Three classic baseball players: SI, *s*M, *s*R	Compare three current baseball stars	
6	Compare two different dogs: ST, SI, *s*M, *s*R	Compare two different parrots	Yes (2)
7	Three women Olympic medalists: ST, SI, D, *s*M, *s*R	Three men with Olympic medals	
8	Two views caffeine, two squirrels, two areas: M	Two views allergy (squirrels, two areas)	Yes (3)
9	Compare two early colonies: ST, M, paper R		Yes (1)
10	Correct others' work (CW) from nine; paper test		
11	Compare dogs & cats: ST, SI, M, R	Chinchillas vs. potbellied pigs	
12	Writing comparative titles: MC		
	Problem-and-solution structure added & integrated with other structures		
13 & 14	I.T. models the structure strategy with problem- &-solution about whales: SI, *s*M, & *s*R		
15	Solutions for troublesome dog: SI, *s*M, *s*R	Solutions for troublesome pig	Yes (2)
16	Question/answer about Washington: SI, *s*M, *s*R	Question & answer about Taft	Yes (2)
17	Fat dog problem-&-solution: ST, SI, *s*M, *s*R	Slimming a plump cat	Yes (2)
18	Rabies problem-&-solution–cats: ST, SI, M, *s*R	Ferrets with rabies & solutions	Yes (2)
19	Problem with cause & comparison of two solutions to eliminate cause: ST, SI, M, *s*R		Yes (1)
20	Heartworm problem & 1 solution: SI, M, & R		
21	Writing with comparison & problem & solution		
22	Review two structures: MC, D, SI, M		
23	Review & author's purpose: SI, ST, MC		
24	Structures building on each other: ST, SI, *s*D, *s*M, R, open-ended questions		
	Cause-and-effect structure added & integrated with other structures		
25	I.T. models cause-and-effect: SI, *s*R		
26	Heat cause changes in eyesight: D, M, R	Heat cause change in skin cancer	Yes (2)
27	Cause-&-effect for informing or persuading		Yes (1)
28	Causal chains–flood: SI, ST, sD, M, R		Yes (1)
29	Cause of dog chaos & milk recipe; SI, ST, M, R	Cause of bird chaos & bun recipe	Yes (4)
30	Complex causes—hawks: SI, ST, *s*D, M, R		Yes (1)
31	Reasons for raining frogs: SI, ST, M, R	Reasons for raining fish	Yes (2)
32	Effects of connected wires: SI, ST, *s*D, M, R		Yes (1)
33	Hailstone effects in India: M, R	Hailstone effects in China	Yes (2)
34	Pony Express: SI, ST, *s*D, M, paper R		Yes (1)
35	Long multiple-structure text: paper R		
36	Review 3 structures: MC, D, SI, M		Yes (8)
37	Writing w/ cause-and-effect + prior structures		
38	Cause-and-effect in 814-word magazine article		
39	Integrating two texts w/ cause-and-effect: M, *s*D		
40	Complex text with all three structures: paper R		

(*Appendices continue*)

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

Table A1 (*continued*)

Standard order	Lesson content in standard ITSS	Available texts for individualized ITSS: Alternative text topic for lesson in adjacent left column	Available texts for individualized ITSS: Easier versions[a]
	Sequence structure added & integrated with other structures[b]		
41	I.T. models sequence–Bill Cody: SI, ST, *s*M, R		
42	Life of George Washington: SI, ST, *s*M, *s*R		
43	Baby teeth timeline: SI, ST, M, R	Kitty teeth timeline	
44	History of PA: SI, ST, M, R	History of state of New York	
45	History of bald eagle: SI, ST, M, R	History of spotted owl	
46	Benjamin Franklin's biography: SI, ST, M, R	Biography of Clara Barton	
47	William Penn biography: SI, ST, M, R	Biography of Frederick Douglass	
48	Michelle Kwan biography: SI, ST, M, R	Biography of Andre Assai	
49	Biography of Wild Bill Hickok: paper R		
50	Writing with sequence + prior structures		
51	Review of all 4 text structures: SI, ST M		
52	Combining structures: SI, ST, M, R		
	Description structure added & integrated with other structures		
53	I.T. models description–swans: SI, ST, M, R		
54	Korean cuisine: SI, ST, M, R	Japanese cuisine	
55	Six Flags Amusement Park: SI, ST, M, R	Disney World	
56	Grizzly bears: SI, ST, M, R	Panda bears	
57	Advertisement –inform/persuade: SI, M, R, MC		
58	Uses for walnuts: SI, ST, D, M, R	Uses for popsicle sticks	
59	Meet poet & painter kid: SI, ST, M, R	Meet runner & quarterback kid	
60	Spending time at the barn: SI, M, R	Spending time around the truck	
61	Rattlesnakes: SI, ST, M, R	Wild turkeys	
62	Tornados: D; Hurricanes: R; then comparing		
63	Writing w/ description + other structures		
64	Review 5 structures: SI, ST, D		
65	Integrating 5 structures: 814-word article		

Note. ITSS = Intelligent Tutoring of the Structure Strategy; I.T. = Intelligent Tutor. Activities engaged in by students in all versions of lessons for each row are indicated with abbreviations in the row's second column: SI = click on signaling word; M = write main idea; R = write full recall; s in italics before M or R indicates scaffolding in the lessons for these tasks; ST = write name of the structure; MC = multiple choice questions; D = filling in or examining a diagram.
[a] Yes indicates for all texts in the row across the table that substantially easier versions (readability below Grade 5) were available for ITSS to assign to students in the individualized condition. [b] Easier versions also were prepared for sequence and description, but insufficient programming resources precluded inclusion.

(*Appendices continue*)

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

Appendix B

Form RP (Rats & Penguins) With Texts and Examples of Scoring

Problem and Solution Text for Prose Recall Task and Scoring Samples

Psychologists who work with rats and mice in experiments often become allergic to these creatures. This is a real hazard for these investigators who must spend hours a week running rats in experiments. These allergies are a reaction to the protein in the urine of these small animals.

At a meeting sponsored by the National Institutes of Health, Andrew J. M. Slovak, a British physician, recommended kindness to rats and mice by the experimenters. Psychologists who pet and talk softly to their rats are less often splattered with urine and the protein that causes the allergic reaction.

Sample recalls and scoring. *"A phitoligist works with rats and mise*." Total ideas called = 7, no structure strategy use (TLS = 2).

"*Well they talked about rats and what rats can do to you and a little about allergies."* Total ideas recalled = 2, no structure strategy use (TLS = 2).

"*Most phycolgist who work with rats become elergic to them."* Total ideas recalled = 10, no structure strategy use (TLS = 2).

"Psycologists who work with rats and expirement of them usually get allergies to them. So when they work with them they get urinated on and start to have trouble with them." Total ideas recalled = 18, no structure strategy use (TLS = 3).

"Scientists who do experiments with mice and rats sometimes get allergic to them, this is why, the protein in the urine causes the allergies.

British scientists observed that if you calm the rat down by talking to it, it won't be scared." Total ideas recalled = 36; structure strategy use (TLS = 6).

"*Scientists who do experiments with rats often have a big problem with allergies. They get allergic to rats because of protein in urine of rats.*

A British doctor at a meeting gave the solution to the problem. The solution is kindness. If you are nice and gentle and pet your rat, the rat won't pee on you." Total ideas recalled = 38, structure strategy use (TLS = 9).

Comparison Text (Italics Indicate Blank for Fill-in Signaling Test)

Emperor penguins and Adelie penguins are *different* from one another. Emperor penguins are large penguins. They are the largest of all penguins and may grow to 4 feet tall. These penguins can weigh more than 90 pounds. Emperor penguins display orange ear patches. They have long, yellow-orange streaked beaks in black faces. Emperor penguins feed principally on shallow water seafood. Emperor penguins live on Antarctica's pack ice.

Unlike the large emperor penguins, Adelie penguins are *smaller* penguins. Adelie penguins grow only about 2 feet high. They weigh only about 11 pounds. Adelie penguins have white ringed, beady, black eyes. Adelie penguins have short, feathered beaks on cute faces. Adelie penguins feed almost entirely on krill. *Same as* the emperor penguins, Adelie penguins live on Antarctica's pack ice.

Sample recall and scoring. *"Emperer penguins are about 4 feet tall and can weigh up to 90 pounds. They feed mostly on shallow water seafood.*

Adalie penguins are much different from Emporer penguins in how big they are and what they eat. They only get about 2 feet tall and weigh up to 11 pounds. Their diet consists of mainly krill. Both types of penguins live on Anarctica's pack ice." Total ideas recalled = 50, use of structure strategy (TLS = 8; comparison of two ideas on at least one common issue with a signaling word cuing the second idea compared—"different" above).

(*Appendices continue*)

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

Appendix C

Table C1
List of Measures With Missing Data (Indicated by Dot)

		Participants with missing data																	
Time and task		s	s	i	i	i	s	s	s	s	s	s	s	i	i	i	s	s	i
Problem-and-solution recall measure																			
Posttest 1																			
Total recall	1				●														
TLS	1				●														
Posttest 2																			
Total recall	1																	●	
TLS	1																	●	
Signaling test measure																			
Posttest 1	2				●	●													
Posttest 2	1																	●	
Comparison recall measure																			
Posttest 1																			
Total recall	4				●						●	●					●		
TLS	4				●						●	●					●		
Posttest 2																			
Total recall	3									●			●					●	
TLS	3									●			●					●	
Achievement goals measure																			
Posttest 1	10	●	●	●			●	●	●	●				●		●			●
Computer attitudes measure																			
Posttest 1	8	●	●					●	●	●				●		●			●
Self-efficacy measure																			
Posttest 1	5	●						●	●					●		●			

Note. Posttest 1 indicates testing directly after ITSS instruction. Posttest 2 indicates testing 1 month after completion of ITSS. ITSS = Intelligent Tutoring of the Structure Strategy; TLS = top-level structure; i = individualized; s = standard.

Received January 22, 2010
Revision received September 21, 2010
Accepted September 21, 2010 ■

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.