McNamara, D.S. & the CSEP Lab. (2006). iSTART: Benefits and Effects of Extended Practice. Technical Report. University of Memphis.

iSTART: Benefits and Effects of Extended Practice

Danielle S. McNamara & the CSEP Lab

University of Memphis

Technical Report

November, 2006

Abstract

Four experiments were conducted to examine the benefits and effects of extended self-explanation practice on students' ability to comprehend texts and to self-explain texts. Extended practice refers to practice on self-explanation and the use of reading strategies after the initial, standard training sessions that include practice on two science texts. The focus across the four experiments was on the type of self-explanation practice and whether the type of practice played a role in furthering and enhancing comprehension and learning from texts. Each of the studies was conducted within a classroom and the teachers were primarily responsible for implementing the interventions. The first experiment compared the effects of teacher-guided and computer-guided (iSTART) practice. The second experiment examined the effects of the students receiving or not receiving feedback during extended feedback training. The third experiment examined the differences between having the students practice on all of the sentences in a text as compared to only target sentences in the text. The fourth experiment examined the effects of training in a remedial class for students with comprehension difficulties.

Acknowledgements: This research was funded by NSF IERI (0241144) and the IES Reading Program (R305G40046). Many people contributed to conducting this study and conducting the following analyses. These include Courtney Bell, Rachel Best, Kyle Dempsey, Laura Holt, Margie Petrowski, , Pavan Pillarisetti, Mike Rowe, Jamie Sayroo, Grant Sinclair, Tenaha O'Reilly, Yasuhiro Ozuru Roger Taylor, and Amy Witherspoon.

iSTART: Benefits and Effects of Extended Practice

Many students have difficulty understanding what they read, particularly the challenging textbooks they encounter in their academic courses (Bowen, 1999; Snow, 2002). Such texts can be better understood by teaching students how to use active reading strategies that enhance comprehension. In this project, we focus on teaching high school just such reading strategies through an automated strategy trainer called iSTART (Interactive Strategy Trainer for Active Reading and Thinking, McNamara, Levinstein & Boonthum, 2004).

To date, numerous experiments assessing the efficacy of iSTART have been conducted with over 1,000 middle school, high school, and college students. The convergence of findings suggests that iSTART is effective in helping students use strategies to learn from texts, and enhances comprehension, particularly among low-knowledge readers (Magliano, Todaro, Millis, Wiemer-Hastings, Kim, & McNamara, 2005; O'Reilly, Sinclair, & McNamara, 2004a, 2004b). The goal of the current project is to examine the benefits of extended practice with iSTART and to examine the use of iSTART by teachers in classrooms. The ultimate goal is to build on our understanding of the use of iSTART in the classroom so that it can be more effectively and efficiently used in high school classrooms.

A Need for Reading Strategy Interventions

The need for reading comprehension interventions is clear from several sources. First, recent research has indicated that students in the United States typically score lower on measures of reading comprehension as compared to students in other countries (Snow, 2002). Even more startling is the statistic that as many as 37% of fourth graders and 26% of eighth graders cannot read at the basic level (NAEP, 2003). In other words, these fourth and eighth graders do not understand what they read. Reading comprehension difficulties are even more pronounced for minorities; NAEP statistics revealed that over half of minority students cannot read at the basic level. Second, other research has indicated that students rarely use reading strategies to help them comprehend text (Pressley & Ghatala, 1990; Rothkopf, 1988; Garner, 1990; Pressley, Wood, Woloshyn, Martin, King, & Menke, 1992), and when they do use strategies, students often implement rudimentary and ineffective methods such as repetition (Garner, 1990). Third, even when students read a text at the basic level, the level of comprehension is typically shallow and lacks the necessary depth for adequate understanding (Best, Rowe, Ozuru, & McNamara, 2005; Langer, 1989; Pressley et al., 1992). In short, there is a strong need for improving reading comprehension among students in the United States. Fortunately, interventions designed to improve comprehension have been successful (e.g., Chi, Bassok, Lewis, Reimann, & Glaser, 1989; Chi, De Leeuw, Chiu, & LaVancher, 1994; Johnson-Glenberg, 2000; McNamara, 2004b; Palincsar & Brown, 1984; Paris, Cross, & Lipson, 1984; Pressley et al., 1992).

iSTART: Automated Strategy Training

In response to the growing need for reading strategy training, McNamara and her colleagues (McNamara, 2004b; McNamara & Scott, 1999) developed a reading strategy training program called iSTART. iSTART is a web-based reading strategy trainer that

provides young adolescent to college-aged students with reading strategy training (McNamara, Levinstein, & Boonthum, 2004). The system uses pedagogical agents to instruct trainees in the use of self-explanation and other active reading strategies to explain the meaning of text while they read. iSTART currently incorporates theoretically motivated Self-Explanation Reading Training (SERT; McNamara, 2004a; McNamara & Scott, 1999), which teaches students to self-explain science texts by using active reading strategies known to facilitate and enhance comprehension, such as paraphrasing, elaborative inferences and bridging inferences (Gernsbacher & Hargreaves, 1988; Pressley et al., 1992; Rosenshine & Meisler, 1994).

iSTART has three modules: *Introduction* (students watch the teacher-agent explain the reading strategies to two student-agents); *Demonstration* (students are quizzed on various aspects of the SERT strategies); and *Practice* (students practice generating typed self-explanations while the program provides feedback on performance). The practice section incorporates feedback that is adaptive to the level of student performance.

The training was motivated by empirical findings that show that students who self-explain text are more successful at solving problems, more likely to generate inferences, construct more coherent mental models, and develop a deeper understanding of the concepts covered in the text. (Chi et al., 1989; Chi et al., 1994). iSTART designed to improve students' ability to self-explain difficult text by combining self-explanation training with metacognitive reading strategy training. Empirical studies with high school and college students have shown that iSTART improves both reading comprehension and the quality of self-explanation during the process of reading (e.g., McNamara, O'Reilly, Best, & Ozuru, 2006; O'Reilly et al., 2004a, 2004b).

Reading Strategies

The intervention coaches students in five reading strategies: comprehension monitoring, paraphrasing, making bridging inferences, predictions, and elaborations. *Comprehension monitoring*, enables the reader to recognize a failure of understanding and it is this recognition that triggers the use of additional active reading strategies. The first such strategy, *paraphrasing*, essentially helps students remember the surface structure of the text by transforming it into more familiar ideas. However, iSTART encourages students to go beyond this basic sentence-focused processing by invoking knowledge-building strategies that link the content of the sentences to other information, either from the text or from the students' prior knowledge. Making *bridging inferences* improves comprehension by linking the current sentence to the material previously covered in the text (e.g., Oakhill, 1984). Such inferences allow the reader to form a more cohesive global representation of the text content (e.g., Kintsch, 1998). Students may also use *prediction* to anticipate the content subsequent text, either by guessing what is coming next or by reminding themselves to watch out for some particular item that will aid comprehension (e.g., Hansen & Pearson, 1983).

Finally, readers may associate the current sentence with their own related prior knowledge using a strategy called *elaboration*. Importantly, readers are encouraged to draw upon logic and common sense, or domain-general knowledge, particularly when they do not have sufficient knowledge about the topic of the text. Research has

established that both domain knowledge and elaborations are associated with improved learning and comprehension (e.g., Pressley et al., 1992; Spilich, Vesonder, Chiesi, & Voss, 1979). Elaboration essentially ensures that the information in the text is linked to information that the reader already knows. These connections to prior knowledge result in a more coherent, and stable representation of the text content (e.g., Kintsch, 1998; McNamara, Kintsch, Songer, & Kintsch, 1996).

CURRENT EXPERIMENTS

These experiments were conducted first in order to examine the benefits and effects of extended self-explanation practice on students' ability to comprehend texts and to self-explain texts. Extended practice refers to practice on self-explanation and the use of reading strategies after the initial, standard training sessions that include practice on two science texts. The focus across the four experiments was on the type of self-explanation practice and whether the type of practice played a role in furthering and enhancing comprehension and learning from texts. Each of the studies was conducted within a classroom and the teachers were primarily responsible for implementing the interventions. The first experiment compared the effects of teacher-guided and computer-guided (iSTART) practice. The second experiment examined the effects of the students receiving or not receiving feedback during extended feedback training. The third experiment examined the differences between having the students practice on all of the sentences in a text as compared to only target sentences in the text. The fourth experiment examined the effects of training in a remedial class for students with comprehension difficulties.

iSTART training was facilitated by the classroom teachers and completed over three consecutive days. Prior to the experiment, the teachers attended a two day workshop that covered general principles of learning, memory, and reading comprehension, as well as a detailed lecture on iSTART and how to utilize it in the classroom. Students worked through the iSTART modules in a sequential order (Introduction, Demonstration and Practice). Although students completed iSTART at their own pace, training commonly required a total of 2.5 to 3 hours. During the initial practice phase, students typed their self-explanations to a brief science text about thunderstorms, followed by another science text about the process of coal formation.

In the section following the description of the experiments and their results, we summarize results of our observations and interviews with the teachers. An important goal of these experiments was to attempt to integrate iSTART into the classroom on a small scale and observe how well it went. For this, we made observations in the classroom, conducted surveys, and held focus groups at the end of the year. This sections reports findings from our surveys and observations. Our goal was to gather teachers' impressions of iSTART and to better understand what they needed to more effectively use iSTART in their classrooms.

Experiment 1: Teacher-guided vs. Computer-guided Extended Practice

This experiment compared the potential instructional impact of providing high school students with extended self-explanation training with either computer-guided practice to teacher-guided practice. In the *teacher-guided* practice, the teacher asked

students to self-explain the texts, and called on several students to provide self-explanations for each sentence of the text. In contrast, for the *computer-guided* practice, the students each used the iSTART system, which provided individualized feedback to their typed self-explanations.

Method

The participants were 78 Shelby County Tennessee high-school students drawn from four 9th grade Biology classes (all taught by the same teacher). All students were given self-explanation training via the iSTART system. However, after this initial training, there was a 6-week period of extended practice training (once per week). Two classes participated in teacher-guided extended practice sessions, whereas two other classes participated in computer-guided extended practice (presented in the same manner as the iSTART Practice module with feedback). Thus, the between-groups factor was the type of extended practice (teacher-guided or computer-guided). There were 45 students in the computer-guided condition and 33 in the teacher-guided condition. However, not all students were present for each of the assessments and thus this number varies somewhat by dependent measure.

Materials

Standardized Reading Skill. General reading skill was measured using a modified version of the standardized Gates-MacGinitie reading skill test for grades 7/9. The test consisted of 48 multiple-choice questions designed to assess student comprehension on several short text passages.

Metacognitive Awareness of Reading Strategies Inventory(MARSI): As part of both the pretest and posttest, the students answered a 30-item questionnaire requiring students to indicate the frequency with which they use particular metacognitive strategies using a five-point Likert scale (i.e., ‘never’ to ‘always’).

Prior Knowledge. Prior knowledge of science was measured with a 20–item, four-alternative, multiple-choice test to assess students’ science knowledge. The test covered several areas including biology, chemistry, earth science, research methods and mathematics. Questions were selected from high school science texts collected from several states (i.e., Colorado, Georgia, Kentucky, Tennessee, Virginia).

Science vocabulary knowledge. Students’ comprehension of scientific vocabulary was measured with a 20-item, four-alternative, multiple-choice test. This test was developed by identifying science terms from the students’ science textbook. The four response options included the target, one conceptually similar answer, one answer from the same area of Biology, and an unrelated answer.

Demographic information. A 26-item demographics questionnaire was used to obtain personal information relevant to reading behavior, such as age, gender and enjoyment of reading. The questionnaire is in a multiple-choice format. This also includes MSI questions.

MSI questions: Metacomprehension Strategy Index (MSI; Schmitt, 1990), a shortened, more robust version of the scale that includes 9 items was used.

Pretest Self-explanation Quality. Student self-explained eight target sentences from a 20-sentence science text. Target sentences were those that the experimenters deemed most critical to understanding the text (e.g., sentences that linked key concepts or contained a large amount of information). The self-explanation tests were administered by a computer program which presented a text to the student one sentence at a time. The designated target sentences were displayed in red font and the non-target sentences were displayed in black font. For the target sentences, the students typed their self-explanations in a designated box on the computer screen. The students were given either a science text on cell repair (Words=335, Sentences=20, Flesch Reading Ease=55.6, Flesch Grade Level=9.5) or seed dormancy (Words=316, Sentences=20, Flesch Reading Ease=57.5, Flesch Grade Level=9.1). Text assignments were counterbalanced such that half the classes self-explained the cell repair text at pretest while the other half self-explained the seed dormancy text.

Pretest Passage Comprehension. Half of the students read a science text on earthquakes while the other half of the students read a text on the origins of the universe. The earthquake passage was a 334 word passage on earthquakes which described the definition and causes of various types of earthquakes. The 23 sentence passage had a Flesch Reading Ease of 56.4 and a Flesch-Kincaid Grade Level of 8.8. The universe passage was a 299-word passage on the origins of the universe, which described theories related to the Big Bang. The 18-sentence passage had a Flesch Reading Ease of 45.7 and a Flesch-Kincaid Grade Level of 10.7. An accompanying set of eight open-ended questions were developed for each passage. The answers to four of the questions could be found within a single sentence of the passage, and are referred to as *text-based* questions. The answers to the remaining four questions required the reader to combine information contained in two or more sentences of the passage, and are referred to as *bridging inference* questions.

Immediate Posttest Self-explanation Quality: Students self-explained eight target sentences from a 20-sentence science text on body temperature (Words=328, Sentences=19, Flesch Reading Ease=48.3, Flesch Grade Level=10.7). Target sentences were those that the experimenters deemed most critical to understanding the text. These sentences were presented with same program used in the pretest self-explanation test.

Immediate Posttest Passage Comprehension. All the students read a science text on heart disease. The heart disease passage was a 307-word passage on the most common types of medical problems involving the heart. The 21-sentence passage had a Flesch Reading Ease of 62.8 and a Flesch-Kincaid Grade Level of 8.1. After reading the text, the students answered 8 open-ended comprehension questions (4 text-based and 4 bridging inference questions).

Delayed Posttest Self-explanation Quality. Students self-explained eight target sentences from a 20-sentence science text. This self-explanation test was administered by the same program used in the pretest. Since counter balancing was used, the half students who self-explained the text on cell repair in pretest, now self-explained a text on seed dormancy, and the other half self-explained the text on cell repair.

Delayed Posttest Passage Comprehension. A data collection problem precluded the use of counter-balancing with the Earthquake/Universe passages. A replacement text on

red blood cells was introduced. The red blood cell passage was a 281 word passage on red blood cells which described the structure and function of red blood cells in the human body. The 20 sentence passage had a Flesch Reading Ease of 56.1 and a Flesch-Kincaid Grade Level of 8.9. Eight open-ended questions (four text-based and four bridging inference questions), similar to those in the pretest, were developed based on the content of the posttest passage.

Procedure

The experiment consisted of four phases: pretest, training, extended practice, and posttest. The pretest, training, and posttest were identical for both conditions. As described above, in the *teacher-guided* practice, the teacher asked students to self-explain the texts, and called on several students to provide self-explanations for each sentence of the text. In contrast, for the *computer-guided* practice, the students each used the iSTART system, which provided individualized feedback to their typed self-explanation.

During the pretest, students were administered the pretest measures in the following order and time frame: Metacognitive awareness of reading strategies inventory (10 minutes), Science vocabulary knowledge (8 minutes), prior science knowledge (10 minutes), Demographics (20 minutes or till the end of the class period), Gates-MacGinitie reading measure (15 minutes), and the passage on either earthquakes or the origins of the universe, along with the set of appropriate comprehension questions (15 minutes), and self-explanation of 8 target sentences in a text on cell repair or seed dormancy (typically 15-30 minutes or till the end of the class period)..

During the training, students progressed through the three sections of the iSTART program over the course of three class periods on three consecutive days. On the following day, the students were administered the immediate posttest tests, which consisted of a comprehensions test on a science text about heart disease, and a self-explanation quality assessment on a text about body temperature.

Beginning the following week, participants entered the extended strategy practice phase. This took place over the course of 6 consecutive weeks. For the extended practice sessions, teachers selected texts (about 20 sentences in length) from their regular textbooks, which they covered in the course curriculum. The passages chosen corresponded to topics taught during the regular class period. Students also took a quiz on the self-explanation texts each week; so, in practice week 1, students self-explained the text on Tuesday and took a quiz regarding the contents of the quiz on Wednesday.

During the delayed posttest, students were administered the posttest measures in the following order and time frame Metacognitive awareness of reading strategies inventory(10 minutes), Gates-MacGinitie reading measure (15 minutes), Demographics (20 minutes or till the end of the class period), a comprehension test on a passage about red blood cells (15 minutes), and a self-explanation quality assessment on a text about either seed dormancy or cell repair (typically 15-30 minutes or end of the class period).

Results

Pretest Equivalence: Reading Comprehension and Prior Science Knowledge. To examine the potential pre-training differences in student ability scores, we performed analyses on

the students' pretest reading skill and prior science knowledge as a function of condition. An analysis of the reading skills scores indicated that there was no effect of condition, $F(1,55) = .49$, $p = .487$. An analysis of the prior knowledge scores indicated that there was no effect of condition, $F(1,55) = .001$, $p = .972$.

Self-Explanation Quality. The quality of students' SE's was evaluated by the iSTART algorithm which rates the quality of SE's on a 0-3 point scale, with 3 being the best score. For each passage, an overall SE quality (SEQ) score was generated by averaging across the 8 target sentences. There were three passages self-explained by the participants. The first, at the beginning of the semester and the last, and the end of the semester, were counterbalanced between themselves. The second was identical for all participants and given right after iSTART training. iSTART training was the same for all participants and thus no differences were expected on this passage. Indeed, average performance on the passage was similar, $F(1,62)<1$, comparing the teacher practice condition ($M=1.81$, $SD=.62$) and the iSTART practice condition ($M=1.83$, $SD=.67$). A mixed-model ANOVA was conducted to examine the differences between the pretest and end-of-the-semester posttest including the within-subjects variable of test and the between-subjects variable of condition. There was a significant change in self-explanation quality (SEQ) from pretest ($M = 0.90$, $SD = .35$) to posttest ($M = 1.17$, $SD = .35$), $F(1,64) = 23.34$, $p <001$. There was no effect of condition, $F<1$, but there was an interaction of test and condition, $F(1,64)=7.56$, $p<.001$, indicating that students in the iSTART condition showed greater gains from pre to post ($M_{pre} = 0.81$, $SD = .38$; $M_{post} = 1.20$, $SD = .26$) than did those in the teacher practice condition ($M_{pre} = 1.03$, $SD = .26$; $M_{post} = 1.13$, $SD = .31$). However, this was due to differences in gains from pretest to posttest; the difference on the delayed posttest was not significant, $F<1$.

In order to examine these results more closely, analyses were done based upon median splits for prior knowledge and reading skill. There were no significant interactions with the individual difference measures.

Immediate Posttest Comprehension. Performance on the text-based ($n = 4$) and bridging inference ($n = 4$) questions was assessed by proportion correct. An ANOVA was conducted including the within-subjects variable of question type and the between-subjects factor of condition (teacher-guided vs. computer-guided). No differences were expected as a function of condition because the manipulation had not yet occurred, $F<1$. There was, as expected, a difference between text-based ($M = .66$, $SD = .28$) and bridging questions ($M = .36$, $SD = .19$), $F(1,71) = 85.77$, $p < .001$. The interaction was not reliable, $F<2$.

Delayed Posttest Comprehension. The original plan in this study was to conduct a pretest posttest comparison on comprehension. However, there was a computer failure and the posttest comprehension tests were replaced by a separate comprehension test on red blood cells. Though the text is matched closely in level of difficulty to the other texts, the performance cannot be compared to pretest performance. Thus, an ANOVA including the within-subjects variable of question type and the between-subjects factor of condition (teacher-guided vs. computer-guided) was conducted. As shown in Table 1, there was a main effect of question type, $F(1,72)=8.44$, $p=.005$. The effect of condition was marginal,

F(1,72)=2.57, p=.113, with an advantage for the teacher-guided condition. The interaction was not significant, F<2.

We further examined whether these effects differed as a function of students' individual differences. When prior knowledge as added as a dichotomous variable in the analysis, the effect of prior knowledge was reliable, F(1,68)=11.22, p<.001 (M_{high}=0.61; M_{low}=0.43), and interestingly, the effect of condition was reliable in this analysis, F(1,68)=5.94, p=.017. There were no interactions, however.

When reading skill was added as a dichotomous variable, there was an effect of reading skill, F(1,65)=6.70, p=.012 (M_{high}=0.61; M_{low}=0.43), and condition, F(1,65)=5.24, p=.025, and a marginal interaction of condition and reading skill, F(1,65)=2.92, p=.092. The advantage for the teacher condition was absent for less skilled readers ($M_{computer}$=0.43; $M_{teacher}$=0.46), F<1. In contrast, the advantage for the teacher-guided practice condition for the skilled readers was quite substantial ($M_{computer}$=0.46; $M_{teacher}$=0.70), F(1,32)=9.91, p=.005.

Table 1.

Proportion correct on the delayed posttest for comprehension

		Mean	SD	N
Text-based	Computer	0.41	0.23	43
	Teacher	0.52	0.28	31
	Total	**0.45**	**0.26**	**74**
Bridging	Computer	0.51	0.27	43
	Teacher	0.57	0.26	31
	Total	**0.53**	**0.26**	**74**

Summary

In summary, this experiment compared the potential instructional impact of providing students with extended self-explanation training with either computer-guided practice or teacher-guided practice. The results showed greater gains from pretest to posttest for those students who practiced using iSTART on the measures of self-explanation quality. There was, in contrast, an advantage for the teacher-guided condition on the comprehension measure. However, this advantage only appeared for the skilled readers. Apparently, the students were greater reading skill were better able to take advantage of the teacher-guided practice sessions. We conjecture that they participated more and contributed more to the group practice sessions.

Our qualitative assessment of this study also led us to the conclusion that teacher-led practice sessions are quite viable and successful. Based on these results, it seems

preferable that the reading strategies were integrated within the classroom and students are provided opportunities to practice using the computer-guided system.

Experiment 2 Feedback vs. No Feedback Extended Practice

This experiment manipulated the use of the iSTART practice module feedback system to assess the relative importance of feedback during extended practice. Students either received feedback in the same manner it was presented in the iSTART practice module during training, or they received no feedback during the extended practice phase (using the practice module of iSTART with the feedback mechanism turned off).

Unfortunately due to extremely high drop out rates, meaningful analyses were not possible and so no findings will be reported here.

Experiment 3: Target Sentences vs. All Sentences Extended Practice

Earlier SERT and iSTART training had students provide self-explanations on *all sentences* of the practice texts. This was hypothesized to be problematic in that it may overemphasize the local (i.e. textbase) features of texts. Also, in extended practice, we were expecting to see fatigue effects from having to self-explain every sentence in a text. Finally, there is a need in the training for students to assess whether self-explanation is necessary or not for comprehension. Moving toward target sentence practice moves them toward assessing the need for self-explanation relative to their level of comprehension. Therefore, an alternative approach was proposed in which students would be trained to provide self-explanations only for target sentences. This led to two sets of predictions.

For textbase comprehension questions, we predicted that there would be: (1) higher scores for students given all-sentence training, and (2) lower scores for students given the target-sentence training. More importantly, for bridging inference comprehension questions, we predicted that there would be: (1) lower scores for students give the all-sentence training and (2) higher scores for students given the target-sentence training.

This final prediction is the focus of this study, namely: that improvement in making correct bridging inferences will be greater for training with target-sentences, compared to training with all-sentences. Thus, in this study, we compared a target-sentence practice condition to an all-sentence practice condition during the extended practice phase. Target sentences were those that the experimenters deemed most critical to understanding the text (e.g., sentences that linked key concepts or contained a large amount of information). For each practice text, approximately one half of the total number of sentences per text deemed most critical to understanding the text were pre-selected as target sentences.

For each week of extended practice, students in the *target sentence condition* read the entire text (using the iSTART practice module with feedback) but were instructed to provide self-explanations only for the target sentences, whereas students in the *all sentence condition* read and provided self-explanations for all sentences in the texts (using the iSTART practice module with feedback).

Method

The participants were 92 Shelby County Tennessee high school students drawn from four 9th grade Physical Science classes (all taught by the same teacher). Two of the classes were assigned to the target-sentence condition and two classes were assigned to the all-sentence condition. Students in the target-sentence and all-sentence conditions were trained with iSTART and participated in the 6-week practice.

Materials

Standardized Reading Skill. General reading skill was measured using a modified version of the standardized Gates-MacGinitie reading skill test for grades 7/9. The test consisted of 40 multiple-choice questions designed to assess student comprehension on several short text passages.

Metacognitive Awareness of Reading Strategies Inventory(MARSI): As a part of pretest and posttest, the students answered a 30-item questionnaire requiring students to indicate the frequency with which they use particular metacognitive strategies using a five-point Likert scale (i.e., 'never' to 'always').

Prior Knowledge. Prior knowledge of science was measured with a 20–item, four-alternative, multiple-choice test to assess students' science knowledge. The test covered several areas including biology, chemistry, earth science, research methods and mathematics. Questions were selected from high school science texts collected from several states (i.e., Colorado, Georgia, Kentucky, Tennessee, Virginia).

Science vocabulary knowledge. Students' comprehension of scientific vocabulary relating to the domain of physical science was measured with a 20-item, four-alternative, multiple-choice test. This test was developed by identifying science terms from the students' science textbook.

Demographic information. A 26-item demographics questionnaire was used to obtain personal information relevant to reading behavior, such as age, gender and enjoyment of reading. The questionnaire was in a multiple-choice format. It also included MSI questions.

MSI questions: Metacomprehension Strategy Index (MSI; Schmitt, 1990), a shortened, more robust version of the scale that includes 9 items was used.

Pretest Passage Comprehension. The pretest passage was a 333-word passage on earthquakes which described the definition and causes of various types of earthquakes. The 25-sentence passage had a Flesch Reading Ease of 61.46 and a Flesch-Kincaid Grade Level of 8.0. An accompanying set of eight open-ended questions were developed for the passage. The answers to four of the questions could be found within a single sentence of the passage, and are referred to as *text-based* questions. The answers to the remaining four questions required the reader to combine information contained in two or more sentences of the passage, and are referred to as *bridging inference* questions.

Pretest Self-explanation Quality. Students self-explained eight target sentences from a 20-sentence science text on seed dormancy. Target sentences were those that the experimenters deemed most critical to understanding the text (e.g., sentences that linked

key concepts or contained a large amount of information). The self-explanation tests were administered by a computer program which presented a text to the student one sentence at a time. The designated target sentences were displayed in red font and the non-target sentences were displayed in black font. For the target sentences, the students typed their self-explanations in a designated box on the computer screen.

Immediate Posttest Passage Comprehension. All the students read the heart disease passage. The heart disease passage was a 307-word passage on the most common types of medical problems involving the heart. The 21-sentence passage had a Flesch Reading Ease of 62.8 and a Flesch-Kincaid Grade Level of 8.1. After reading it, the students then answered 8 open-ended comprehension questions relating to each passage (4 text-based and 4 bridging inference questions).

Immediate Posttest Self-explanation: Students self-explained eight target sentences from a 20-sentence science text on body temperature control. Target sentences were those that the experimenters deemed most critical to understanding the text. These sentences were presented with the same program used in the self-explanation pretest.

Delayed Posttest Passage Comprehension. The posttest passage was a 477-word passage on the origins of the universe, which described theories related to the Big Bang. The 31-sentence passage had a Flesch Reading Ease of 39.1 and a Flesch-Kincaid Grade Level of 11.5. As with the pretest passage, an accompanying set of eight open-ended questions were developed for the passage. The answers to four of the questions could be found within a single sentence of the passage, and are referred to as *text-based* questions. The answers to the remaining four questions required the reader to combine information contained in two or more sentences of the passage, and are referred to as *bridging inference* questions.

Delayed Posttest Self-explanation Quality: Students self-explained 8 target sentences from a 20-sentence science text on cell repair. Target sentences in this text were chosen using the same method as those chosen for the pretest. The posttest was also administered via the same program that was used in the pretest.

Procedure

The experiment consisted of four phases: pretest, training, immediate posttest, extended practice, and posttest. The pretest, training, and posttest were identical for both conditions. As described above, Students in the *target sentence* condition read the entire text (using the iSTART practice module with feedback) but were instructed to provide self-explanations only for the target sentences (identified by highlighting), whereas students in the *all sentence* condition read and provided self-explanations for all of the sentences in the texts (also using the iSTART practice module with feedback).

During the pretest, students were administered the pretest measures in the following order and time frame: Metacognitive awareness of reading strategies inventory (10 minutes), Science vocabulary knowledge (8 minutes), prior science knowledge (10 minutes), Demographics (20 minutes or till the end of the class period), Gates-MacGinitie reading measure (15 minutes), a comprehension test on a science text about earthquakes (15 minutes), and self-explanation quality assessment on a science text about seed dormancy (typically 15-30 minutes or till the end of the class period).

During the training, students progressed through the three sections of the iSTART program over the course of three class periods on three consecutive days. On the following day, the students were administered the immediate posttest tests, which consisted of a comprehension test on a science passage about heart disease and a self-explanation quality assessment with a text about body temperature control.

Beginning the following week, the participants entered the extended strategy practice phase. This took place over the course of six consecutive weeks. For the extended practice sessions, the teachers selected texts (about 20 sentences in length) from their regular textbooks, which they covered in the course curriculum. The passages chosen corresponded to topics taught during the regular class period. Students also took a quiz on the self-explanation texts each week; so, in practice week 1, students self-explained the text on Tuesday and took a quiz regarding the contents of the quiz on Wednesday. During the posttest, students were administered the posttest measures in the following order and time frame: Metacognitive awareness of reading strategies inventory (10 minutes), Gates-MacGinitie reading measure (15 minutes), Demographics (20 minutes or till the end of the class period), a comprehension text on a passage about red blood cells (15 minutes), and a self-explanation quality assessment on text on cell repair (typically 15-30 minutes or end of the class period).

Results

Pretest Differences. To examine the potential pre-training differences in student ability scores, we performed analyses on the students' pretest reading skill and prior science knowledge as a function of condition. An analysis of the reading skills scores indicated that there were no differences between conditions, $F(1,81) < 1$. An analysis of the prior knowledge scores indicated that were no differences between conditions, $F(1,87) < 1$.

There was, however, a marginal difference between groups on the pretest comprehension scores. A 2x2 ANOVA revealed a significant difference between text-based (M=0.65, SD=0.26) and bridging (M=0.26, SD=0.20) questions, $F(1,90) = 163.16, p < .001$, and a marginal advantage for students assigned to the all-sentence condition (M=0.49) in comparison to the target-sentence condition (M=0.42), $F(1,90)=3.39$, $p=.069$. There was no interaction. Because texts were not counterbalanced, pretest/posttest differences are confounded by potential differences in difficulty between the texts. Thus, analyses of posttest comprehension are conducted both with and without the overall pretest comprehension scores included as a covariate. The analyses are reported without the covariate unless the covariate alters the results.

Self-Explanation Quality. For each passage, an overall SE quality (SEQ) score was generated by averaging across the 8 target sentences. The quality of students' SE's was evaluated by the iSTART algorithm which rates the quality of SE's on a 0-3 point scale, with 3 being the best score.

Table 2.

Self-explanation scores (from iSTART) as function of condition and test

		Self-Explanation Quality		
Test	Condition	Mean	SD	N
Pretest	All-Sentence	0.92	0.33	43
	Target-Sentence	0.79	0.42	43
Immediate Posttest	All-Sentence	1.85	0.47	41
	Target-Sentence	1.63	0.49	40
Delayed Posttest	All-Sentence	1.29	0.40	39
	Target-Sentence	1.22	0.41	46

Each test should be analyzed separately because the texts were not counterbalanced. Nevertheless, it is worth noting that there was substantial improvement in scores from the pretest to the immediate posttest, $F(1,77)=173.17$, $p<.001$, and from the pretest to the delayed posttest, $F(1,80)=60.03$, $p<.001$. Neither of this effects depended on experimental condition or individual differences.

On the pretest, the differences between conditions was marginal, $F(1,84)=2.78$, $p=.099$. Thus there was a hint at a prior advantage for the students in the all-sentence condition. On the immediate posttest, notably before any manipulation had occurred, this difference widened and became significant, $F(1,71)=4.41$, $p=.039$. On the delayed posttest, where we expected differences, there were no differences at all, $F<1$. None of these results depended on individual differences.

Based on this data, there seems to be no difference in effects of target and all-sentence practice. There is an apparent increase in performance as a function of training. However, it also appears that performance decreases between the immediate posttest and the delayed posttest. Indeed, the data would be perfectly sensible if we'd have somehow reversed the two posttests. Alas that is not the case. Nonetheless, we can't really say that there is indeed a reduction in performance from the immediate to the delayed posttest because different texts were use and because these analyses are making use of the automated scores from iSTART. We'll be able to make stronger conclusions from this data when we also have human scores for the data.

Comprehension

Immediate Posttest. Analyses of posttest comprehension are conducted both with and without the overall pretest comprehension scores included as a covariate. The analyses are reported without the covariate unless the covariate alters the results.

A 2x2 ANOVA revealed a significant difference between text-based ($M=0.66$, $SD=0.25$) and bridging ($M=0.40$, $SD=0.19$) questions, $F(1,81)= 71.16$, $p<.001$. Neither the main effect of condition ($F<1$) nor the interaction ($F<2$) were significant. An effect of

condition was not expected because the immediate posttest is after iSTART training, which all of the participants received, and before extended practice where the manipulation occurs.

Delayed Posttest. A 2x2 ANOVA revealed a significant difference between text-based (M=0.66, SD=0.25) and bridging (M=0.40, SD=0.19) questions, $F(1,81)= 71.16$, $p<.001$. There was no effect of condition, $F<1$. An interaction of condition and question type was expected: the all-sentence condition was expected to have an advantage at the textbase level, compared to an advantage on bridging questions for the target-sentence condition. This interaction, shown in Figure 1, was non-significant, $F(1,84)= 2.52$, $p=.116$, without the pretest comprehension covariate, but marginal with the covariate, $F(1,83)=3.31$, $p=.073$. Nonetheless, the effect of condition was not reliable for either the text-based or the bridging inference questions. Separate analyses by high and low knowledge, however, revealed that the effect of condition on bridging inference questions was reliable for high knowledge students (when pretest comprehension was included as a covariate), $F(1,41)=4.27$, $p=.045$. The effect indicated that for high-knowledge students, the target-sentence condition resulted in better performance on the bridging inference questions (M=0.36) than did the all-sentence condition (M=0.23). This difference did not occur for low knowledge students ($F=0$; both M = 0.26). The difference on the text-based questions was not reliable regardless of individual differences.

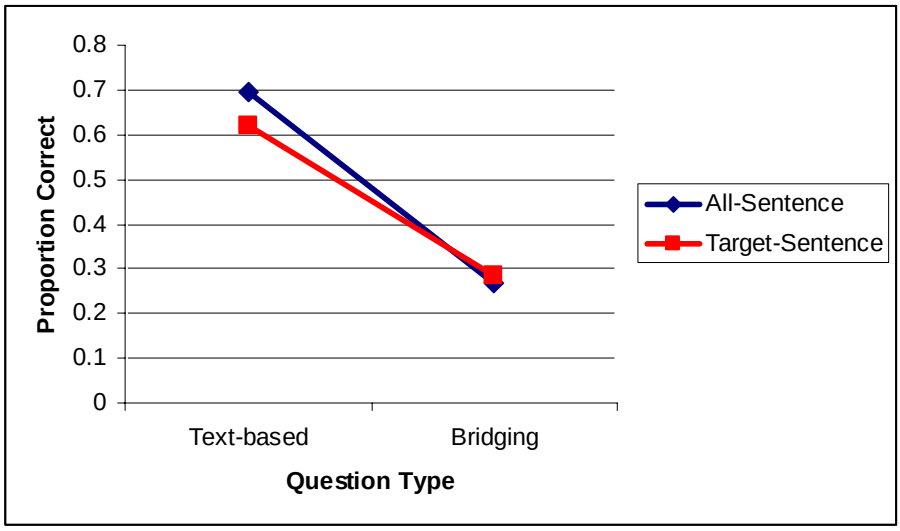

Figure 1. Proportion correct on the delayed posttest as a function of question type and condition

There was some support for the key prediction of the study, such that there will be greater improvement in bridging inference comprehension scores when students are trained using target-sentences instead of being trained to self-explain all of the text sentences. However, this effect only occurred for the high-knowledge readers. This result is not surprising because the reader needs to have sufficient knowledge to generate bridging inferences. The target-sentence training apparently helped them to learn how to more effectively use that knowledge.

Experiment 4: Remedial Class – Matched Extended Practice

A brief description of this study can be found in Taylor, O'Reilly, Sinclair, & McNamara (2006). This study was conducted primarily because the administration at a local high school approached our research team for some assistance in helping with a remedial reading program. The administration had identified the 50 students exhibiting the lowest scores on an eighth grade reading comprehension proficiency exam, and offered them the opportunity to participate in a remedial reading class to enhance their reading comprehension skills. When being asked to conduct our study, the plan was to include all 50 students. However, the school's finances only allowed one course per semester to be offered. Thus, only 25 students were offered the class in the fall semester. Fifteen students accepted the offer.

Our goal was to integrate iSTART into the remedial reading class. At the beginning of the academic year, we designed a curriculum that included iSTART training, followed by three weeks of extended practice using the iSTART reading strategies. We compared the iSTART-trained students to regular students on several comprehension measures at various points throughout the training/extended practice period to assess potential benefits of the iSTART training.

All parts of the intervention (i.e. pre-testing, iSTART training and extended practice, post-testing) were conducted by the reading teacher who taught the remedial class and all of the texts used during the extended practice sessions were taken directly from the students' textbooks.

Method

The initial sample consisted of 93 ninth grade students from a suburban Tennessee high school who were enrolled in a ninth grade life science course. Fifteen of these students, identified by the school as having reading comprehension difficulties based on an eighth grade reading exam, were enrolled in a special reading class designed to improve their reading comprehension. Three of the 15 reading class students did not complete the pretest reading ability measure and were, therefore, not included in these analyses. The remaining twelve students served as the *iSTART* group. Twelve students from the life science class were found who matched the iSTART students based on their pretest scores on a standardized measure of reading ability. This matched group of 12 students from the life science class served as the *control* group.

Materials

Standardized Reading Ability. Reading ability was measured by the Gates-MacGinitie reading skill test for grades 7/9 (form K at pretest, form L at posttest). The Gates-MacGinitie is a 48 item, four-alternative, multiple-choice measure of reading skill designed to assess comprehension on several short text passages.

Metacognitive Awareness of Reading Strategies Inventory(MARSI): As a part of pretest and posttest, the students answered a 30-item questionnaire requiring students to indicate the frequency with which they use particular metacognitive strategies using a five-point Likert scale (i.e., 'never' to 'always').

Prior Knowledge. Prior knowledge of science was measured with a 20–item, four-alternative, multiple-choice test to assess students' science knowledge. The test covered several areas including biology, chemistry, earth science, research methods and mathematics. Questions were selected from high school science texts collected from several states (i.e., Colorado, Georgia, Kentucky, Tennessee, Virginia).

Science vocabulary knowledge. Students' comprehension of scientific vocabulary was measured with a 20-item, four-alternative, multiple-choice test. This test was developed by identifying science terms from the students' science textbook. The four response options include the target, a conceptually similar answer, an answer from the same area of Biology, and an unrelated answer.

Demographic information. A 26-item demographics questionnaire was used to obtain personal information relevant to reading behavior, such as age, gender and enjoyment of reading. The questionnaire is in a multiple-choice format. This also includes MSI questions.

MSI questions: Metacomprehension Strategy Index (MSI; Schmitt, 1990), a shortened, more robust version of the scale that includes 9 items was used.

Pretest Passage Comprehension. The pretest passage was a 333-word passage on earthquakes which described the definition and causes of various types of earthquakes. The 25-sentence passage had a Flesch Reading Ease of 61.46 and a Flesch-Kincaid Grade Level of 8.0. An accompanying set of eight open-ended questions were developed for the passage. The answers to four of the questions could be found within a single sentence of the passage, and are referred to as *text-based* questions. The answers to the remaining four questions required the reader to combine information contained in two or more sentences of the passage, and are referred to as *bridging inference* questions.

Pretes/Posttest Self-explanation Quality. Student self-explained eight target sentences from a 20-sentence science text on seed dormancy at pretest, and eight target sentences from a 20-sentence science text on cell repair at posttest. Target sentences were those that the experimenters deemed most critical to understanding the text (e.g., sentences that linked key concepts or contained a large amount of information). The self-explanation tests were administered via a computer program which presented a text to the student one sentence at a time. The designated target sentences were displayed in red font and the non-target sentences were displayed in black font. For the target sentences, the students typed their self-explanations in a designated box on the computer screen.

Immediate Posttest - Passage Comprehension. All of the students read the heart disease passage. The heart disease passage was a 307-word passage on the most common types of medical problems involving the heart. The 21-sentence passage had a Flesch Reading Ease of 62.8 and a Flesch-Kincaid Grade Level of 8.1. After reading it, the students then answered 8 open-ended comprehension questions relating to each passage (4 text-based and 4 bridging inference questions). Only iSTART participants completed this test; thus, this measure is not included in the analyses.

Self-explanation Quality. Student self-explained eight target sentences from a 20-sentence science text on body temperature. Target sentences were those that the experimenters deemed most critical to understanding the text. These sentences were

presented with same program used in the pretest self-explanation test. Only iSTART participants completed this test; thus, this measure is not included in the analyses.

Delayed Posttest Passage Comprehension. The posttest passage was a 284-word passage on red blood cells which described the structure and function of red blood cells in the human body. The 21-sentence passage had a Flesch Reading Ease of 59.06 and a Flesch-Kincaid Grade Level of 8.4. Eight open-ended questions (four text-based and four bridging inference questions), similar to those in the pretest, were developed based on the content of the posttest passage. An example of a text-based question was "How does sickle-cell disease get its name?". An example of a bridging inference question was "Explain why blood plasma is a poor carrier of oxygen?"

Delayed Posttest Self-explanation Quality: Target sentences in this text were chosen similarly as in pretest and also the test administered by the same program used in the pretest.

Procedure

The experiment consisted of four phases: pretest, training, extended practice, and posttest. The pretest and posttest phases were identical for both the iSTART and control conditions, while the iSTART training and extended practice were performed only by students in the iSTART condition.

During the pretest, students in the iSTART and control group were administered the pretest measures in the following order and time frame: Metacognitive awareness of reading strategies inventory (MARSI) (10 minutes), Science vocabulary knowledge (8 minutes), Prior Knowledge of Science (10 minutes), Demographics (20 minutes or till the end of the class period), Gates-MacGinitie reading measure (Form K, 15 minutes). This was followed by a passage on earthquakes, along with the set of appropriate comprehension questions (15 minutes), and a self-explanation quality assessment with a text on seed dormancy (typically 15-30 minutes or end of the class period).

During the training, students in the iSTART group progressed through the three sections of the iSTART program over the course of three class periods on three consecutive days.

The immediate posttest occurred on the after the last day of training. It consisted of an assessment of comprehension with a science passage on heart disease, and a measure of self-explanation quality with a science passage on body temperature control (with 8 target sentences). The heart disease text was not presented to the control participants and thus no results are provided on this measure.

The extended strategy practice phase began the week after the immediate posttest. Participants in the iSTART condition entered the extended strategy practice phase. This phase took place over the course of three consecutive weeks. Each week the students self-explained two texts that were taken directly from the textbooks used in their regular courses. On day one, the students used the computer portion of iSTART practice to self-explain text one, with feedback being provided by the pedagogical agent in the same manner as during the practice phase of iSTART training. On the following day (day two), the teacher led the class in a self-explanation task of text 1, and provided feedback on

performance. On day three, students used the practice section of iSTART to self-explain text 2, while on the following day (day four) the teacher led the practice for text 2. The time spent on practice each week on the two texts was approximately three hours, one and a half hours of iSTART practice and one and a half hours of teacher-guided practice. On the fifth day of the week, students were given quizzes on the material that had been covered during that week. The quizzes were also administered to the control subjects, who had studied the material in the classroom in the manner normally used by the life science teacher.

The delayed posttest began the week following the extended practice. This consisted of measures given in the following order and time frame: Metacognitive awareness of reading strategies inventory (10 min), Gates-MacGinitie reading measure (Form L, 15 minutes), Demographics (20 minutes or till the end of the class period), a comprehension test using a passage on red blood cells (15 minutes), and a self-explanation quality assessment using a text on either seed dormancy or cell repair (typically 15-30 minutes or end of the class period).

Results

Pretest Equivalence: Reading Comprehension and Prior Science Knowledge

The 12 students in the iSTART condition were matched with 12 students in the control condition based on the number of correct responses and the number of questions attempted on the standardized measure of reading comprehension (Gates-MacGinitie pretest scores). T-tests revealed no significant difference in pretest reading scores between the iSTART condition (M = 9.75, SD = 4.29) and the control condition (M = 9.92, SD = 4.36), $t(22) = 0.09$, $p = .93$, and no significant difference in pretest prior science knowledge scores between the iSTART condition (M = 0.35, SD = 0.09) and the control condition (M = 0.39, SD = 0.10), $t(22) = 0.97$, $p = .34$.

A computer problem prevented 6 of the 12 iSTART participants from completing the pretest comprehension test (Earthquake Text) and the computer problems likely disrupted performance for the other 6 participants. The data collected indicated that there was a significant effect of question type, $F(1,16)=10.45$, $p=.005$, largely because they couldn't answer any of the bridging questions (M=.01); thought, higher, the students also performed quite poorly on the text-based questions (M=0.16). There was no differences between the two groups, $F<1$, and no interaction, $F<1$.

Delayed Posttest

Self-Explanation Quality. For each passage, an overall SE quality (SEQ) score was generated by averaging across the 8 target sentences. The quality of students' SE's was evaluated by the iSTART algorithm which rates the quality of SE's on a 0-3 point scale, with 3 being the best score. A mixed-model ANOVA revealed that there was no significant change in self-explanation quality (SEQ) from pretest (M_{iSTART} = 1.30, SD = .57; $M_{control}$ = 1.53, SD = .48) to posttest (M_{iSTART} = 1.09, SD = .27; $M_{control}$ = 1.39, SD = .38), $F(1,17) = 2.68$, $p = .120$. There was no significant effect of condition, $F(1,17) = 2.41$, $p = .139$, and also no significant interaction between time (pretest, posttest) and condition (iSTART vs. control), $F(1,17) = 0.11$, $p = .743$. Thus, in contrast to the previous studies, there was no improvement and no effect of training.

Comprehension. The students read a text about red blood cells at the end of the extended training period, approximately 3 months following the beginning of the experiment. Proportion correct as function of question type and condition is presented in Table 3.

A mixed ANOVA was conducted including the within-subjects variable of question type and the between-subjects variable of condition. The effect of question type was not significant, $F<1$. There was a marginal effect of condition, $F(1,22)=3.77$, $p=.065$, and a marginal interaction of condition and question type, $F(1,10)=2.97$, $p=.099$. The interaction reflects the observation that the effect of condition was not reliable for the text-based questions, $F<2$, but significant for the bridging inference questions, $F(1,22)=8.66$, $p=.008$.

Table 3.

Proportion correct on delayed posttest comprehension as function of question type and condition

	Condition	Mean	SD	N
Text-based	Control	0.27	0.13	12
	iSTART	0.32	0.24	12
	Total	0.30	0.19	24
Bridging	Control	0.18	0.15	12
	iSTART	0.35	0.15	12
	Total	0.27	0.17	24

Thus, the key finding from this work was that this type of instructional intervention resulted in significantly greater performance on comprehension questions that required students to make inferences between different parts of a science text passage. This finding is particularly important because this improved performance occurred on a more challenging aspect of comprehension.

One of the main strengths of this study was its high ecological validity. Upon receiving a request to integrate the iSTART system into a local public high-school's remedial reading program, we were able to provide the teacher with the ability to conduct the training and practice in a normal classroom setting using texts from the student's textbooks. In the past, iSTART, and its feedback algorithms that are implemented during the practice phase, only included two training texts. This study indicated that iSTART practice can be readily extended to other texts.

A second encouraging factor is that, even with a limited number of students in a single remedial reading class, it is clear that the iSTART training had a significant facilitative impact on reading comprehension ability. While computer problems at pretest

resulted in partial data loss, it is still clear that in comparison to the control group, the iSTART group performed significantly better on the posttest comprehension questions. Moreover, this result emerged on the more challenging bridging inference questions. In other words, the iSTART training helped these students, identified as needing remediation by the school system, to more deeply understand difficult expository science texts. We are currently performing more fine-grained analyses, including the hand-coding of the self-explanations, to more precisely determine the mechanism via which the training can enhance learning.

Previous research has suggested that iSTART training is effective because it improves the quality of students' elaborations (McNamara et al., 2006) and encourages students to use their general knowledge to construct deeper representations of the text (McNamara, 2004a). The current results make sense with respect to these previous findings given that the more challenging bridging inference questions often require the use of elaborations and general knowledge.

TEACHERS' IMPRESSIONS OF iSTART

An important goal of these experiments was to attempt to integrate iSTART into the classroom on a small scale and observe how well it went. For this, we made observations in the classroom, conducted surveys, and held focus groups at the end of the year. This sections reports findings from our surveys. Our objective was to gather teachers' impressions of iSTART and to better understand what they needed to more effectively use iSTART in their classrooms.

Teachers' views of an iSTART workshop

One of the key goals of our project is to integrate iSTART into the high-school classroom. To meet this goal, it is necessary to train teachers to administer the training and determine how our current system complies with the current needs and constraints of the classroom. We have taken initial steps toward integration by training teachers to administer reading strategy training using iSTART. During the summer of 2004, eight teachers from two Shelby County Tennessee high schools participated in a three-day iSTART training workshop. During the workshop the teachers 1) were presented with the theoretical basis for the development of iSTART (including sessions on the cognitive processes involved in text comprehension and on models of text processing and how these related to the reading strategies taught with iSTART), 2) went through the iSTART v2.0 strategy trainer, and 3) provided feedback on parts 1 and 2. The eight participants included three science teachers (who participated in the Experiments 1-3), a reading-class teacher (Experiment 4), two English teachers, one high-school reading lab director, and one high-school computer technician (and former teacher). At the end of the three-day workshop, we administered a questionnaire to each teacher. The overarching goals of the questionnaire were to gain ideas about 1) the level of background instruction about text processing and about the computer system that we need to provide to teachers in order to implement iSTART in regular high-school classrooms, and 2) the perceived usefulness of iSTART as a classroom teaching tool. All questions required teachers to provide open-ended responses. Answers were classified into categories (see tables below).

For questions concerning the workshop, we classified the teachers' responses to each question as either positive (i.e., "yes" response) or negative based on the content. Table 4 reports the frequencies of positive responses produced by the eight teachers. The data indicated that overall, teachers felt positive about the training workshop. All teachers found the workshop helpful and most of them indicated that it helped them understand iSTART. Importantly, the data indicated that teachers had some prior knowledge regarding the reading strategies used in iSTART (e.g., concept of elaboration) and used strategy training in their classroom activities. Teachers' knowledge about the strategies and the fact that they already apply some of the iSTART strategies is likely to make it easier to integrate iSTART into their curriculum.

Table 4.
Opinions regarding the reading strategy workshop

Question	Positive responses N
Was the presentation clear and comprehensible?	8
Did you find the information useful to your learning about and understanding iSTART?	6 (2 teachers did not answer question)
Did you find the information beneficial/useful beyond how it might be applied to iSTART training?	7 (1 teacher did not answer question)
Have you been exposed to any of the presented concepts previously? If yes, which concepts?	8
Do you already use/apply any of the presented concepts when teaching? If yes, which concepts?*	7
Do you think you would use the presented information in some non-iSTART capacity in your classes?	7 (1 teacher did not answer the question)

Note that the teachers did not specify which strategies they used

Table 5 shows responses to questions about the usefulness of iSTART. All of the teachers indicated that the iSTART training workshop helped them understand how and why iSTART facilitates text comprehension. All but one teacher foresaw using iSTART after the workshop intervention (the reading lab director provided the negative responses and works with students exhibiting word decoding deficiencies, a lower-level processing deficiency that cannot be addressed with iSTART training). Of the teachers who specified how they would implement iSTART, approximately half reported that they would use it with students of all levels, whereas half indicated they would use it for "lower-level" students (e.g., students with comprehension difficulties). Further, the five teachers who answered the question about frequency of use reported that they would implement self-explanation practice on a regular basis. Overall, the workshop data are encouraging with regard to the teachers' perceptions on using iSTART in the classroom.

Table 5.
Perceived use of iSTART

Question	Responses N
Did you find the iSTART training left you with a clear understanding of how and why iSTART can facilitate text comprehension?	8 = positive response
Do you foresee using iSTART in the classroom after the IIS intervention at your school?	7 = positive response 1 = negative response
In what context might you use iSTART in the classroom (i.e., all students, students with comprehension difficulties, etc.)?	3 = all students 3 = lower-level students 1 = train teachers to use 1 = not applicable
How often would you use iSTART and/or its related concepts (i.e., training only, training and extended practice, once a week, daily, etc.)?	4 = weekly practice 1 = monthly practice 2 = did not specify 1 = not applicable

Observations of teachers administering iSTART and extended practice sessions

We observed teachers administering iSTART and extended practice sessions in high school science classes (Experiment 1-3) and the reading class (Experiment 4). The purpose of the observations was to examine whether and how teachers can administer the training and the extended practice sessions. The teachers who participated in the 2004 summer workshop guided students through the training and extended practice sessions.

Two types of observation were conducted: 1) University of Memphis iSTART researchers recorded observations of the teachers administering iSTART training and extended practice, and 2) the teachers recorded their observations of the students' experiences of iSTART and extended practice. The observation data were collected from four teachers. Teacher A conducted Experiment 1, Teacher B conducted Experiment 2, Teacher C conducted Experiment 3, and Teacher D conducted Experiment 4. The following results comprise information collected by all teachers. Results for individual teachers are reported when there is a point of interest specific to individual teachers.

Difficulties using iSTART. To determine the ease or difficulty with which teachers use iSTART in the classroom, we analyzed the number and types of questions teachers asked iSTART researchers while administering the iSTART training. There were two iSTART researchers present for each training session; a computer expert for iSTART, and a postdoctoral psychologist specializing in reading strategy training. We recorded the frequency with which teachers asked questions about: 1) computer operation (i.e., "How does the student log on?"), 2) instructions/procedures for proceeding through iSTART (i.e., "What exactly is the character-agent asking the student to do here?"), 3) contents of the texts used in iSTART (i.e., "What does the word/paragraph mean?"), and 4) technical difficulties or other questions (i.e., "What do I do if the program crashes?"). These

frequencies were recorded for each of the three iSTART modules (Introduction, Demonstration, Practice).

Table 6 shows frequency of questions for each type of question for each module. The most commonly asked questions were computer-related. More specifically, most of the questions related to log-in problems (e.g., "How do I log a student on to the system?") and technical problems related to issues such as frozen screens and "dead" batteries (e.g., "Can you change the battery in the computer?"). Very few questions were asked concerning the instructions/procedures to be followed during training or about the meaning of the content of the texts used.

Table 6.

Questions asked by teachers while administering the iSTART training

Question	Introduction	Demonstration	Practice	Total
Computer operation	18	12	6	36
Instruction /procedure	0	1	0	1
Text content	0	0	6	6
Technical difficulties	16	12	9	37
Other	3	0	0	3

Observations of iSTART training. The second analysis focused on students' motivation for and enjoyment of the iSTART training, the difficulties encountered by students during training, and the difficulties encountered by teachers while guiding students through training. Observations were made by both the iSTART researchers and the teachers at the end of each training session using the following scale:

0 = No/none
1 = A little/few
2 = Quite/some
3 = For the most part
4 = Extremely/many

Table 7 shows the mean ratings from the iSTART researchers and the teachers (based on the 0-4 scale) regarding student motivation, enjoyment and difficulties. The results indicate that the iSTART researchers and teachers gave very similar ratings for students' motivation and enjoyment and also student and teacher difficulties. Whereas

students' motivation and enjoyment were slightly above average, student and teacher difficulties were quite low.

Table 7.

Mean ratings of student motivation and difficulties with iSTART

Topic	iSTART team ratings Mean (SD)	Teacher ratings Mean (SD)
Student motivation	2.8 (0.5)	2.5 (0.8)
Student enjoyment	2.3 (0.6)	2.4 (0.8)
Student difficulties	0.3 (0.6)	0.7 (1.0)
Teacher difficulties	0.3 (0.6)	0.3 (1.0)

Observations of extended practice

As with the training observations, two types of observation were conducted during the extended practice sessions: 1) observations by the University of Memphis iSTART researchers of teachers conducting the practice sessions; and 2) teachers' observations of the students' participation in the extended practice sessions.

The observations are based on the four teachers who participated in Experiments 1 - 4. Foremost, we compared teacher-guided practice to computer-based practice to establish the ways in which the two methods of practice are able to supplement iSTART. Hence, the data were divided into two categories: teacher-guided practice (the teacher-guided condition of Experiment 1 and the teacher-guided sessions of Experiment 4) and computer-based practice (all other conditions and sessions using the iSTART v2.0 Practice module).

Observations were made for all practice sessions for Experiment 4, but only selected practice sessions were observed in the other experiments due to practical constraints. That is, for Experiments 1-3, we observed 2 classes in each experiment each week. For example, for Teacher A, Experiment 1 (who conducted live and computer sessions), we observed one teacher-guided and one computer-based practice session each week. We alternated the specific classes observed each week so that data was recorded for students in all classes across the extended practice period.

iSTART researcher's observations of extended practice

The first analysis reports the observations by the iSTART researchers of the teachers. The purpose of the analysis was to evaluate the types of information that teachers provide to students during practice sessions. The analysis of the teacher-guided training also provided an opportunity to investigate whether teachers themselves understand the strategies and are able to teach them effectively.

We used a checklist measure to rate the extent to which teachers provided effective instruction (e.g., instruction about the strategies included in iSTART). The following scale was used to record the teaching methods used.

(0) **Not observed:** Strategy was never observed

(1) **Rarely:** Receives isolated use and/or little time in the class/clearly not prevalent/emphasized component of teaching and learning across the class.

(2) **Occasionally:** Receives minimal or modest time or emphasis in the class. Not a prevalent/emphasized component of teaching and learning in the class.

(3) **Frequently:** Receives substantive time or emphasis in the class. A prevalent component of teaching and learning in the class.

(4) **Extensively:** Receives substantive time and/or emphasis in the class. A **highly** prevalent component of teaching and learning in the class.

Table 8 reports the mean score (based on the 0-4 coding system outlined above) for each teaching method used in the teacher-guided and computer-based extended practice sessions. The data indicate that teachers provided extensive instruction about self-explanation and the reading strategies during the teacher-guided practice sessions, but fairly infrequently for the computer practice sessions. Also, during teacher-guided practice, teachers focus more on content teaching, such as explaining word meanings, sentence meanings, and global text understandings (probably due to the interactive nature). Notes made by the iSTART researchers regarding the quality of the teacher-guided sessions and computer sessions indicate that during the teacher-guided sessions, teachers continuously talked about the strategies (e.g., explained meanings of paraphrasing and elaboration, and provided content-related information). Overall, teacher-guided practice sessions seemed to offer an opportunity to deliver more instruction about strategies and topic contents than the computer practice, which may be important for bolstering effects of training.

Thus far, our data indicate that teachers were able to effectively conduct self-explanation practice sessions; that is, describe and offer feedback about the iSTART strategies. However, notes made by the iSTART researchers also indicated that teachers' input during teacher-guided practice was not necessarily correct and did not necessarily adhere to our teaching protocol. For instance, Teacher A was reported to provide incorrect information about the meaning of comprehension monitoring. Also, Teacher A sometimes focused on self-explaining texts at the paragraph-level as opposed to the sentence-level taught in iSTART. As indicated above, the protocol was for the teacher to follow the sentence-by-sentence method, but in general the teacher focused on paragraphs. Note, however, that is not necessarily a bad approach and may improve global understanding more so than explanation at the sentence level. Indeed, it is a modification we are incorporating into future versions of iSTART.

Table 8.

Mean score for methods used in the extended practice sessions

Teaching method	Teacher-guided practice Mean (SD)	Computer practice Mean (SD)
Self-explanation instruction	3.0 (1.4)	1.8 (1.7)
iSTART strategies instruction	3.2 (1.2)	1.3 (1.4)
Other reading strategies instruction	0.8 (0.9)	0.1 (0.6)
Word meanings	1.5 (0.9)	0.1 (0.3)
Sentence meanings	1.5 (1.6)	0.5 (1.3)
Whole text meaning	1.3 (1.8)	0.1 (0.3)
Teacher feedback	3.6 (0.9)	0.06 (0.3)
Peer feedback	0.5 (0.6)	0

Teacher observations of extended practice sessions

This analysis focused on teacher ratings of students' in the teacher-guided and computer-based practice sessions. Following the training protocol and using the following scale, teachers rated student motivation and enjoyment, difficulties encountered by students,) and difficulties encountered by teachers.

0 = No/none
1 = A little/few
2 = Quite/some
3 = For the most part
4 = Extremely/many

Table 9 shows the mean rating (based on the 0-4 scale) regarding student motivation, enjoyment and difficulties in the teacher-guided and practice sessions. The findings indicate that while ratings of motivation and difficulties were similar across the two types of practice sessions, levels of enjoyment were higher for the teacher-guided practice than for the computer-based practice.

To expand on the levels of enjoyment finding, notes taken by the iSTART researchers were recorded. The notes indicate that in the teacher-guided sessions, students were highly likely to participate and be engaged in the self-explanation activity. The main criticism relating to the teacher-guided sessions was that some students struggled to produce self-explanations orally, or felt social pressure in the presence of peers. On the other hand, there were reports of students becoming bored with the computer sessions and sometimes going "off task."

Table 9.

Mean scores for teacher ratings

Dimension	Teacher-led practice Mean (SD)	Computer practice Mean (SD)
Student motivation	2.8 (0.8)	2.5 (0.6)
Student enjoyment	3.1 (0.8)	2.4 (0.7)
Student difficulties	1.1 (1.2)	0.8 (0.9)
Teacher difficulties	0.9 (1.2)	0.3 (1.0)

CONCLUSIONS

We conducted four experiments in high school classrooms to examine how well iSTART would be received in the classrooms and to answer several questions regarding the delivery of extended training. We've seen the effects of reading strategy training wane over time and thus hoped that extended practice would lead to better retention. In many cases, we've also observed effects of training to only occur at the textbase level of understanding. So, a second goal was to examine whether extended training would help students to learn to generate more and better inferences, and thus to construct better situation model level understandings of text.

In Experiment 1, we saw that iSTART extended practice led to greater improvement on self-explanation quality, but that teacher-led practice led to better comprehension performance for the skilled readers. Apparently, the students were greater reading skill were better able to take advantage of the teacher-guided practice sessions. Indeed, our observations were that some students struggled to produce self-explanations orally, and felt social pressure in the presence of peers. Nonetheless, the teacher-guided practice sessions seemed to involve more instruction about strategies and topic contents than the computer practice, and students were more likely to participate and be engaged in the self-explanation activity. Students became bored with the repetitive computer sessions. These factors may be critical for bolstering effects of training. On the other hand, the advantages for teacher-led practice were isolated to the skilled readers, these results point to the need to have a synergy of both computer and teacher-led practice in extended training.

In Experiment 3, we examined the effects of having either target sentences or all sentences during extended practice sessions. We hoped that having the students practice with target sentences would help students to learn to make more bridging inferences, and it did, but only for the high-knowledge students. The low knowledge students didn't possess the requisite domain knowledge to generate the inferences. Nonetheless, the results confirm that providing target sentences to self-explain rather than forcing students to self-explain all of the sentences potentially leads to a greater tendency to make lings between sentences. Also important, the target sentence condition – which is less grueling

for the students – did not lead to any loss in gains from training in comparison to the all-sentence condition. Our final study with the remedial comprehension class students further showed that extended training can lead to benefits in performance at the situation model level. In that study, we observed the effects of training only on the bridging inference questions.

Thus, overall, the results are promising. They indicate that students given more opportunities to practice self explanation and reading strategies are more likely to learn how to build a situation model level understanding of difficult text. However, extended practice is not 'fun' for the students. The initial training is quite novel, but listening to the animated agents repeatedly for weeks in a row is perhaps a bit annoying to them. These results point to the need to incorporate opportunities to practice the strategies during classroom activities.

References

Best, R. M., Rowe, M., Ozuru, Y., & McNamara, D. S. (2005). Deep-level comprehension of science texts: The role of the reader and the text. *Topics in Language Disorders, 25,* 65-83.

Bowen, B. A. (1999). Four puzzles in adult literacy: Reflections on the national adult literacy survey. *Journal of Adolescent and Adult Literacy, 42,* 314-323.

Chi, M. T. H., Bassok, M., Lewis, M. W., Reimann, P., & Glaser, R. (1989). Self-explanations: How students study and use examples in learning to solve problems. *Cognitive Science, 13,* 145-182.

Chi, M. T. H., De Leeuw, N., Chiu, M., & LaVancher, C. (1994). Eliciting self-explanations improves understanding. *Cognitive Science, 18,* 439-477.

Garner, R. (1990). When children and adults do not use learning strategies: Toward a theory of settings. *Review of Educational Psychology, 60,* 517-529.

Gernsbacher, M. A., & Hargreaves, D. J. (1988). Accessing sentence participants: The advantage of first mention. *Journal of Memory and Language, 27,* 699–717.

Hansen, J., & Pearson, P. D. (1983). An instructional study: Improving the inferential comprehension of good and poor fourth-grade readers. *Journal of Educational Psychology, 75,* 821-829.

Johnson-Glenberg, M. C. (2000). Training reading comprehension in adequate decoders/poor comprehenders: Verbal versus visual strategies. *Journal of Educational Psychology, 92,* 772-782.

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

Langer, E. (1989). *Mindfulness.* Reading, MA: Addison-Wesley.

Magliano, J. P., Todaro, S., Millis, K. K., Wiemer-Hastings, K., Kim, H. J., & McNamara, D. S., (2005). Changes in reading strategies as a function of reading training: A comparison of live and computerized training. *Journal of Educational Computing Research, 32,* 185-208.

McNamara, D. S. (2004a). SERT: Self-explanation reading training. *Discourse Processes, 38,* 1-30.

McNamara, D. S. (2004b). Aprender del texto: Efectos de la estructura textual y las estrategias del lector. *Revista Signos, 37,* 19-30.

McNamara, D. S., Kintsch, E., Songer, N. B., & Kintsch, W. (1996). Are good texts always better? Interactions of text coherence, background knowledge, and levels of understanding in learning from text. *Cognition and Instruction, 14,* 1-43.

McNamara, D. S., Levinstein, I. B., & Boonthum, C. (2004). iSTART: Interactive strategy training for active reading and thinking. *Behavioral Research Methods, Instruments, and Computers, 36,* 222 - 233.

McNamara, D. S., O'Reilly, T., Best, R. & Ozuru, Y. (2006). Improving adolescent students' reading comprehension with iSTART. *Journal of Educational Computing Research. 34,* 147-171.

McNamara, D. S., & Scott, J. L. (1999). Training reading strategies. *Proceedings of the Twenty-First Annual Meeting of the Cognitive Science Society,* Hillsdale, NJ: Erlbaum.

Oakhill, J. (1984). Inferential and memory skills in children's comprehension of stories. *British Journal of Educational Psychology, 54,* 31-39.

O'Reilly, T., Sinclair, G. P., & McNamara, D. S. (2004a). iSTART: A web-based reading strategy intervention that improves students' science comprehension. In D. Kinshuk, G. Sampson, & P. Isaías (Eds.), *Proceedings of the IADIS International Conference Cognition and Exploratory Learning in Digital Age: CELDA 2004* (pp. 173-180). Lisbon, Portugal: IADIS Press.

O'Reilly, T. P., Sinclair, G. P., & McNamara, D. S. (2004b). Reading strategy training: Automated versus live. In K. Forbus, D. Gentner & T. Regier (Eds.), *Proceedings of the 26th Annual Cognitive Science Society* (pp. 1059-1064). Mahwah, NJ: Erlbaum.

Palincsar, A. S., & Brown, A. L. (1984). Reciprocal teaching of comprehension-fostering and monitoring activities. *Cognition and Instruction, 2,* 117-175.

Paris, S., Cross, D., & Lipson, M. (1984). Informed strategies for learning: A program to improve children's reading awareness and comprehension. *Journal of Educational Psychology, 76,* 1239-1252. Pressley, M., & Ghatala, E. (1990). Self-regulated learning: Monitoring learning from text. *Educational Psychologist, 25,* 19-33.

Pressley, M., Wood, E., Woloshyn, V., Martin, V., King, A., & Menke, D. (1992). Encouraging mindful use of prior knowledge: Attempting to construct explanatory answers facilitates learning. *Educational Psychology, 27,* 91-109.

Rosenshine, B., & Meister, C. (1994). Reciprocal teaching: A review of the research. *Review of Educational Research, 64,* 479-530.

Rothkopf, E. (1988). Perspectives on study skills training in a realistic instructional economy. In C. E. Weinstein, E. Goetz, & P. A. Alexander (Eds.), *Learning and study*

strategies: Issues in assessment, instruction, and evaluation (pp. 275-286). San Diego: Academic.

Schmitt, M. (1990). A questionnaire to measure children's awareness of strategic reading process. *The Reading Teacher, 49,* 454-461.

Snow, C. (2002). *Reading for understanding: Toward an R&D program in reading comprehension.* Santa Monica, CA: RAND.

Spilich, G., Vesonder, G., Chiesi, H., & Voss, J. (1979). Text processing of domain-related information for individuals with high and low domain knowledge. *Journal of Verbal Learning and Verbal Behavior, 18,* 275-290.

Taylor, R. S., O'Reilly, T., Sinclair, G. P., & McNamara, D. S. (2006). Enhancing learning of expository science texts in a remedial reading classroom via iSTART. *Proceedings of the 7th International Conference of the Learning Sciences* (pp. 765-770). Mahwah, NJ: Erlbaum.