

FINDINGS

This section describes the principal results of our reading strategy training experiments conducted using SERT (experimenter-delivered training; McNamara, in press) and iSTART (automated web-based training; McNamara, Levinstein, & Boonthum, 2003). The general structure of all the experiments is reported in the activities section. Only unique details are reported here, such as the type of school, the subject population, or particular experimental procedures.

SERT: EXPERIMENTER-DELIVERED READING STRATEGY TRAINING

We carried out experimenter-delivered strategy interventions in three schools; a suburban school in which many students were upper to middle SES; an inner-city school in which most students were African-American and from lower to middle SES, and a school that was primarily white, located in a rural Appalachian community (please see the 2002 progress report for comparisons between schools). Here we report some of the principal findings from a subset of the experiments that were conducted. We concentrate on the experiments conducted in the lower aptitude schools (i.e., the inner city and rural schools).

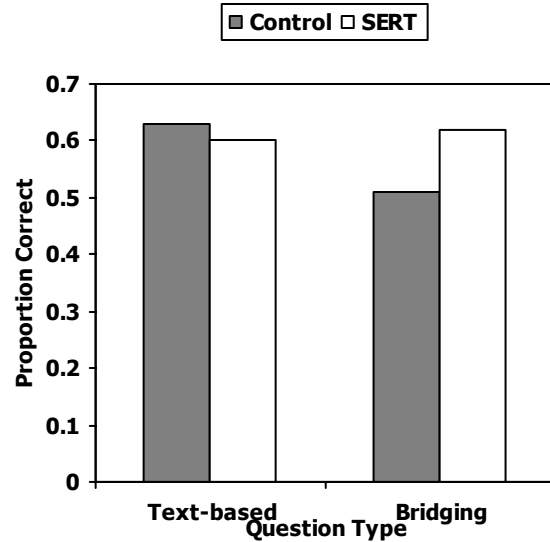
The goal of these experiments was to examine the effectiveness of the three reading strategy interventions described in the activities section for 9th to 12th graders. The first analysis is on 175 chemistry class students from the inner city and rural Appalachian schools. The eight participating chemistry classes were randomly assigned to the SERT, Previewing, INSERT, or control conditions. The second analysis includes 78 biology and earth science honors class students from the inner-city school. The four classes were randomly assigned to SERT or control conditions. The third analysis includes 153 mainstream biology class students from the inner city school. The nine classes were randomly assigned to SERT, Preview, or control conditions.

Experiment 1: Chemistry Classes

Reading comprehension accuracy on the Virus text (presented immediately after training) was analyzed as a function of training condition (SERT, INSERT, Previewing or control) and question type (text-based, bridging). There was a reliable effect of condition, $F(3,112) = 5.6$, $p < .001$, which reflected better comprehension for SERT ($M = 0.60$) and Control ($M = 0.60$) conditions than the Previewing ($M = 0.48$) and INSERT ($M = 0.52$) conditions. One explanation for the control condition advantage is that the participants in the control condition may have used strategies for which they have had more practice, whereas those in the experimental conditions were asked to use strategies for which they had little practice. Since the students in the chemistry classes tend to be high performers, the well-practiced strategies may have been as effective as SERT. Alternatively, it is possible that not all of the participants in the SERT condition used the SERT strategies while reading. Our future analyses will look at what strategies the students reported using when they read the text, and whether that factor influenced comprehension (i.e., this data was collected from the participants but has not yet been entered into the database).

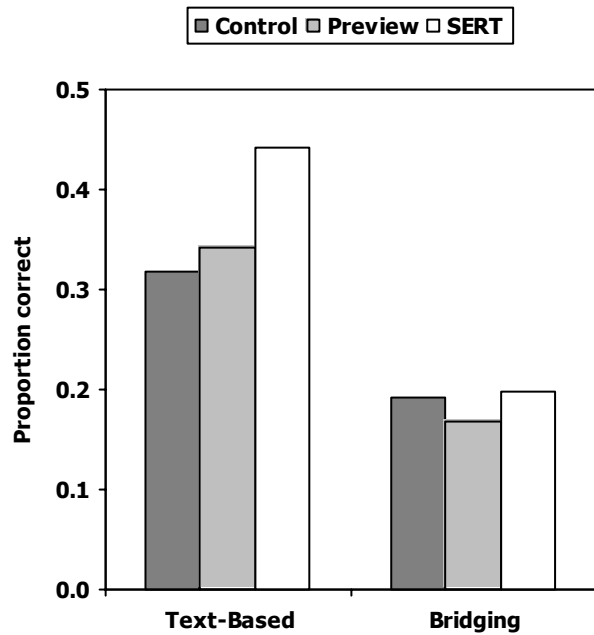
Experiment 2: Honors Biology and Earth Science Classes

Reading comprehension accuracy on the Virus text (presented immediately after training) was analyzed as a function of training condition (SERT, control) and question type (text-based, bridging). There was a significant interaction between question type and training condition, $F(1,74)= 7.8$, $p < .01$. As shown in the figure to the right, this interaction reflected a reliable difference between SERT and control conditions on the bridging questions, $F(1,74)=3.8$, $p < .05$. There was no effect of condition for text-based questions. Individual differences in knowledge or reading skill did not interact or modify these results. Hence, for these relatively high-performing honors students, SERT resulted in reliable gains for the most difficult questions. Moreover, the students in the SERT condition show no differences between the text-based and bridging questions.



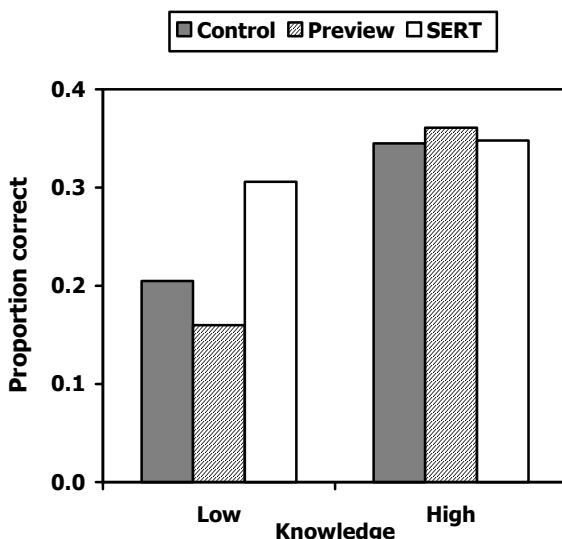
Experiment 3: Mainstream Biology Students

This section reports on the analysis performed on comprehension accuracy for the mainstream biology class students from the inner-city school ($N=153$). These students are among the lowest in terms of reading skill and knowledge aptitudes according to the pretest assessments. Our analyses indicated that there were no effects of training immediately after training for the Virus text. Our second set of analyses center on students' comprehension performance on open-ended questions a week after training when they were asked to read a text about earthquakes (results for the multiple choice questions are comparable to these results). First, a 3x2 mixed model ANOVA including condition (SERT, Preview, and control) and question type (text-based, bridging) showed a reliable interaction of condition and question type, $F(2,150) = 4.73$, $p < .01$. As shown in figure to the right, there was no effect of training on the bridging inference questions compared to 10-12% gains for the SERT condition on the text-based questions. Thus, these inner-city mainstream biology students benefited from SERT in terms of better understanding the explicit content in the text.



Further analyses revealed that these gains were more pronounced for the students with less-knowledge about science (135 of the 153 students were administered the prior knowledge test). A 3x2x2 mixed model ANOVA including the condition, prior knowledge (high, low), and question type yielded a significant effect of prior knowledge, $F(1,129) = 25.68$, $p < .01$ ($M_{\text{highK}} = 0.35$, $M_{\text{lowK}} = 0.22$), and an interaction of knowledge and condition, $F(2,129) = 3.25$, $p < .05$. As shown in the figure below, there was no effect of condition for the high-knowledge students compared to 10-15% gains for the low-knowledge SERT students. These gains were apparent for both question types, but were most pronounced for the text-based questions.

Low-knowledge mainstream biology students benefited more from SERT than did high-knowledge students



Summary

We are particularly encouraged by the results discussed above because these students are in the most experimentally challenging environments of the three schools with which we worked. For example, attendance at the inner-city school was very low (resulting in high attrition rates) and behavioral problems were of greater concern than in the other two schools. Quite frankly, we were not expecting substantial outcomes as a function of our efforts. To find, on the contrary, that SERT provided comprehension benefits for the neediest of the children with whom we worked is highly encouraging.

We also see that the more skilled students (e.g., the honors students) gain from SERT at deeper levels of comprehension as assessed by the bridging inference questions. In contrast, low knowledge students gain most at the textbase level of comprehension. These results mirror those that we have observed using iSTART to deliver training. Hence, this trend is discussed further in the conclusion section.

iSTART: COMPUTERIZED TRAINING

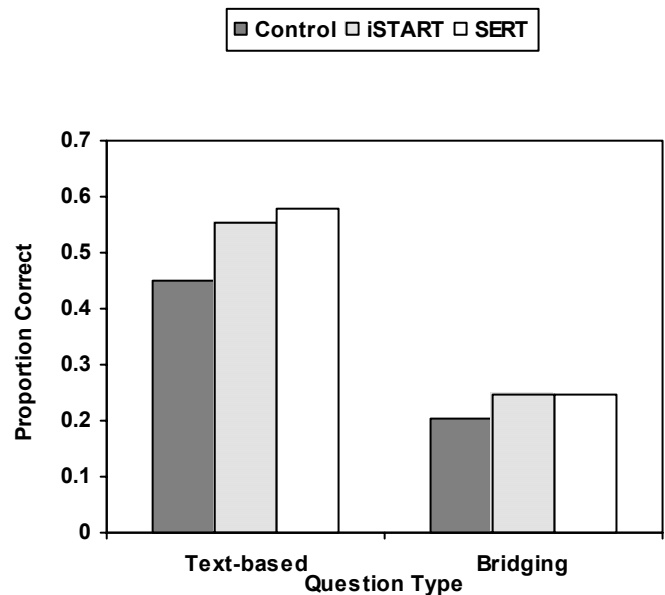
Experiment 4: Is iSTART Comparable to SERT?

The purpose of our first experiment with iSTART v1.0 was to examine whether iSTART was as effective at improving comprehension as the live SERT training among a sample of college students (O'Reilly, Sinclair, & McNamara, 2002). The 281 students were enrolled in introductory biology course and their laboratory sections were assigned to one of three conditions: live SERT (trained by an instructor), iSTART (trained by the computer program), or a control condition.

After training, the students read a text about cell division (see McNamara, 2001). Before answering the open-ended questions to assess comprehension, they were asked what strategies they had used while reading the text. Whereas 75% of iSTART and SERT participants reported using at least one of the reading strategies taught by iSTART or SERT, only 3% of control participants reported using these higher level strategies. This result indicates that the iSTART and SERT participants became more aware of active reading strategies as a function of the reading strategy training.

Second, we compared the comprehension performance of control participants with SERT and iSTART participants who explicitly stated that they had used SERT strategies. Our focus here is on whether training (SERT or iSTART) affects comprehension for students who indicated use of the strategies learned during the training. One concern with this selection criterion is that elimination of those students who did not claim to use the strategy may have resulted in selectively eliminating low performing students from the iSTART and SERT condition but not from the control condition. However, our analyses confirmed that there were no differences between these groups in terms of prior knowledge or reading abilities.

A 3x2 mixed ANOVA including condition and question type confirmed a reliable effect of question type, $F(1,28)= 3.9$ $p<.05$ ($M_{\text{text-based}} = 0.53$; $M_{\text{bridging}} = 0.23$). There was also a main effect of condition, $F(2,2)= 5.7$, $p<.05$, indicating that iSTART ($M = 0.40$) and SERT ($M = 0.41$) participants answered more questions correctly than control participants ($M = 0.33$). However, the main effect was qualified by a significant interaction between the training condition and question type, $F(2,24)= 3.21$, $p<.05$. As shown in the figure to the right, the SERT and iSTART participants answered significantly more text-based questions correctly than did control participants. This advantage was not apparent on the bridging inference questions. These findings mirror the results reported by McNamara in two previous studies (2001; in press).



In both of these previous studies using the same text with college students, greater active processing resulted in gains on the text-based measures. We expect that this result arises because the text is very challenging (i.e., low cohesion and a difficult topic).

Experiment 5: Do the Effects of iSTART Vary as a Function of Reading Skill?

In this study, 44 college student participants produced self-explanations before and after iSTART v1.0 training. Reading skill was assessed via the Nelson Denny Test of reading skill. A 2x2x2 mixed model ANOVA including reading skill (skilled, less skilled), test (pre, post), and question type (text-based, bridging) on open-ended question accuracy was conducted. A 3-way interaction, $F(1,43) = 4.21, p < .5$, indicated that less skilled readers performed better on text-based questions after training ($M = .66$) than before ($M = .54$), whereas skilled readers performed better on bridging questions after training ($M = .49$) than before ($M = .39$). This result mirrors the effects found for SERT found for high-schools students reported earlier in this report.

We assessed reading strategies exhibited in the self-explanation protocols using a 3-level system: (1) sentence-focused strategy: focus on only the current sentence, (2) local strategy: provide limited elaboration of the sentence, or relate it to the prior sentence, (3) global reading strategy: establish how the current sentence is linked to the theme of the text. The protocol analyses confirmed that the students' self-explanations focus on the text versus the situation model varied as a function of reading skill. Skilled readers produced fewer self explanation containing a sentence-focused strategy after training than before, whereas less skilled readers did not changes with respect to this strategy ($F(1,43) = 3.49, p = .069$). With respect to local strategies, there was a trend in the data such that both skilled and less skilled readers produced more self-explanation containing local bridges after training than before ($F(1,43) = 3.00, p = .10$). Finally, skilled readers produced more global self explanation before and after training than did less skilled readers ($F(1,43) = 5.87, p < .05$). Less skilled readers struggle to understand the explicit content of difficult texts and training helps them to learn strategies to compensate for this difficulty. On the other hand, skilled readers focus on strategies that lead to a more coherent situation model after training presumably because they have the skills necessary to construct a textbase representation prior to training.

Experiment 6: Evaluation of iSTART with Middle School Students

Our second experiment with iSTART v1.0 examined the effectiveness of iSTART for middle-school students. In this study, we investigated the quality of self-explanations produced by students' and their ability to comprehend a science passage as a function of iSTART training.

The sample consisted of 42 eighth and ninth grade children enrolled in a learning program, called Learning Bridge, designed to motivate high-achieving students from under-privileged backgrounds. The students were assigned to either the experimental group (trained using iSTART) or the control group on a random basis. However, following post-testing, the control students were trained with iSTART so that they were provided with the benefits of training. Overall, we have training data for all 42 subjects as well as data examining differences in trained and untrained students.

Self-explanation Protocol Analyses

We have scored (using the procedure described in the activities section) the self-explanations produced for a 13-sentence passage about the stages of thunderstorm development by the 42 students during the practice phase of iSTART (Best, Ozuru, & McNamara, 2003). We center here on the effects of individual differences and sentence difficulty on paraphrasing and elaboration. In general, students often used paraphrasing (in 91% of the self-explanations) and these paraphrases were often successful (94%). Paraphrases tend to be more accurate in easy than difficult (70% vs. 53%) sentences, $F(1, 39) = 19.5$, $p < .001$. Paraphrases that used fewer words from the sentence (called distant paraphrases) occurred more frequently for high-knowledge than low-knowledge students (63% vs. 44%), $F(1, 39) = 7.5$, $p < .01$.

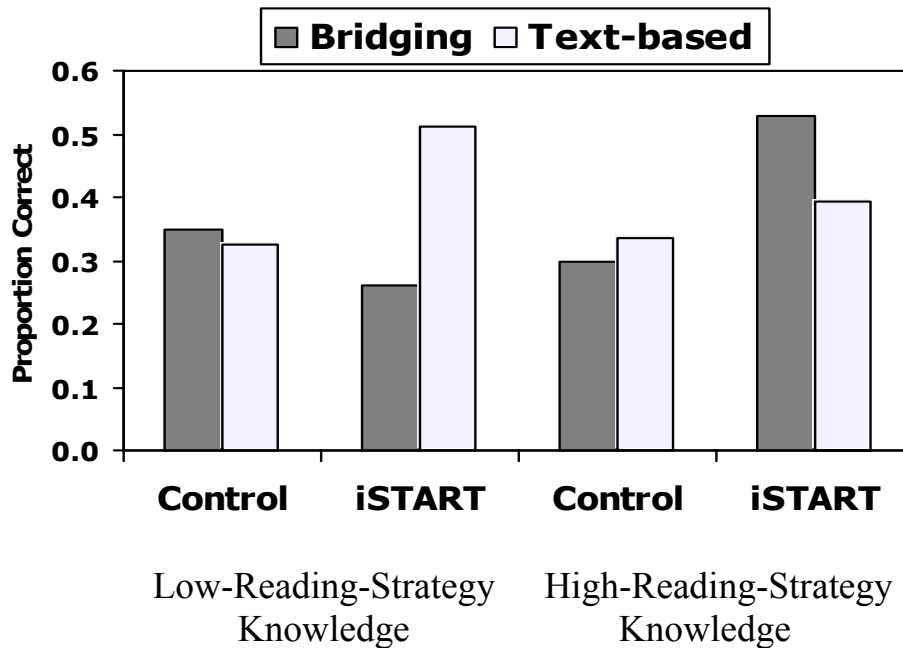
Elaborations were less frequent than paraphrases, occurring in 41% of the self-explanations. More elaborations were produced by skilled than less skilled readers (49% vs. 30%), $F(1,33) = 6.1$, $p < .05$. And, skilled readers produced more accurate elaborations than less skilled readers (59% vs. 49%), $F(1,39) = 9.2$, $p < .01$. The number of elaborations and their accuracy also depended on the difficulty of the sentence. More elaborations were generated for easy sentences than for difficult sentences (43% vs. 35%), $F(1,33) = 3.4$, $p = .07$, and the elaborations for easy sentences were more likely to be accurate than those for difficult sentences (72% vs. 49%), $F(1, 39) = 8.6$, $p < 0.05$.

Of all elaborations produced, 66% were based on general knowledge/common sense, 23% were based on prior text, and 11% were based on scientific knowledge. The majority of the elaborations (68%) focused on the current sentence, whereas only 11% focused on a global level of comprehension beyond the current sentence (knowledge building). However, skilled readers were more likely than less skilled readers to generate knowledge building elaborations, $F(1,33) = 7.6$, $p < 0.01$.

In sum, skilled readers produced higher quality elaborations (their elaborations contained more knowledge building, sentence linking, and were more accurate). However, these high quality elaborations were often produced for “easy sentences.” Knowledge level was found to be important for the production of distant paraphrases. These results suggest that readers’ ability to generate successful self explanations depends on the match between sentence difficulty and the readers’ skill and knowledge levels.

Post training Reading Comprehension Measure

The effect of iSTART training on students’ reading comprehension was assessed with a text about heart disease (modified from McNamara, Kintsch, Songer, & Kintsch, 1996). As depicted in the figure below, there was significant 3-way interaction between question type, prior knowledge about reading strategies, and condition $F(1,34) = 6.26$, $p < .017$). For low-reading-strategy participants, iSTART outperformed controls on text-based questions, $t(18) = -2.26$, $p = .037$. In contrast, for high-reading-strategy participants, iSTART resulted in better performance on bridging questions, $t(18) = -2.26$, $p = .037$.



Experiment 7: Evaluation of the Effects of Feedback during Training

One question regarding iSTART regards whether feedback should be general or specific in terms of the participant's strategy use. Hence, we conducted an experiment with 78 college students to examine the effects of feedback on the quality of self-explanations. The students were first given abbreviated self-explanation training, and then practiced self-explanation on a computer. The computer was a simplified web-based version of the iSTART practice module. It was simplified by excluding animated agents, and it only presented a text box with the passage, a box for the student to type in the self-explanations, and a box where feedback was provided to the student. The simplifications were intended to magnify the effect of feedback. Participants were randomly assigned to one of four conditions. In a no-feedback condition, the participants did not receive any feedback on their self-explanations. In a general-only condition, they only received feedback consisting of one word (e.g., "OK") based on the general quality of their self-explanation. In a complete condition, they received general feedback and also feedback on using particular strategies, if they were identified. An example of full feedback would be: "Good. I'm pleased to see that you are bridging here." The computer randomly picked one out of six responses for each strategy. In a random condition, the computer gave full feedback based on randomly generated cosines. The tutor either gave no, general (e.g., "Good."), complete (e.g., "Good. I see that you are bridging."), or random feedback after each sentence.

Using LSA-based algorithms (see activities section), the self-explanations were assigned scores (sentence-focused = 1; local = 2; global = 3) and categorized as containing a paraphrase or a bridging inference. There was a main effect of condition on scores, $F(3, 70) = 2.53, p < .05$. The mean scores for no-feedback, random, general-only, and complete conditions were 1.74, 1.91, 2.25, and 2.23, respectively. Post-hoc comparisons indicated

that the no-feedback conditions received significantly lower quality scores than the general-only and full conditions, but not the random condition. In regard to specific reading strategies, as the feedback increased, the frequency of paraphrases decreased whereas the frequencies of bridging inferences increased, as indicated by a strategy by condition interaction, $F(6, 140) = 2.58, p < .05$. In regard to the accuracy of identification, the computer assignment of general reading strategy (i.e., sentence-focused, local, global) was highly correlated with human judgments ($r = .79$). This study suggests that students' self-explanations improve with appropriate feedback. Accordingly, one change made to Version 2.0 of iSTART was to increase encouragement to use more elaborations (text-based or knowledge-based) when the use of paraphrases is detected.

Experiment 8: Evaluation of iSTART v2.0

Our final experiment investigates the efficiency with which the modified iSTART program (see Section b for details on the modification) teaches high-school students to use the SERT strategies. This study is being conducted in the students' classrooms. One outcome thus far from this experiment has evolved from the opportunity to overcome some potential technological challenges before we move toward the integration of iSTART into classroom (i.e., implemented by the teacher). The primary technological challenges thus far were ensuring that we had internet connections and somewhat unexpectedly having to use a variety of operating systems. The main experimental goals are to assess the quality of self-explanations and science text comprehension as a function of iSTART v2.0 training both immediately following training and after a delay.

The participants include 467 Shelby County Tennessee high-school students. The classes, taught by 4 teachers, include 10 biology classes (10th grade) and 10 physical science classes (9th grade). All of the students were administered a pretest aptitudes assessment in September, 2003, and will be administered a posttest in April, 2004. Half of the classes were trained in November, 2003, and the remaining half will be trained in January, 2004. Using matched-samples assignment (based reading skill and prior knowledge), half of the students in each class were assigned to the iSTART condition and half were assigned to a control condition.

The students in the control condition are taught to design web pages containing science-related information. The web design task was developed to engage students with a computerized task, while at the same time being exposed to scientific content. This controls for the novelty of computer use as well as exposure to scientific information. Following the iSTART and web design training phase, both the experimental group and control group are required to complete a self-explanation task and a science comprehension task to examine whether iSTART training facilitated the production of better quality self-explanations and enhanced the comprehension of science texts.

The training schedule (i.e., this does not include the beginning-of-year pretest and end-of-year posttest schedules) for the two conditions was carried out over a four-day period, which is outlined in Table 1.

Table 1: Experimental schedule

Day	Experimental group	Control group
1	iSTART introduction	Instruction on web design
2	iSTART demonstration	Web design demonstration
3	iSTART practice (thunderstorm text)	Web design practice
4	iSTART practice (coal text)	Web design practice
5	Immediate posttest assessments	Immediate posttest assessments

We are in the process of running the study and thus data are not available for this experiment. We are currently processing data collected in the pretest and the training data collected in November, 2003.

SUMMARY

Our results across these studies indicate that SERT and iSTART produce comparable improvements in self-explanation quality and comprehension of science text. These gains seem to occur at the student's zone of proximal development. That is, the student shows improvement at the lowest level of comprehension that can benefit from greater active processing. For less skilled, or low-knowledge students, improvement occurs at the textbase level of comprehension. These gains seem to arise from explanations that are more text-based, rather than elaborative explanations that go beyond the text. Skilled, or high-knowledge readers are more likely to gain at the situation model level of comprehension as assessed in bridging inference questions. And, these skilled students' self-explanations will tend to become more knowledge-building and less focused on the sentence after training. However, these patterns depend on the difficulty of the text. We expect that these patterns will emerge when the text difficulty is moderately high. However, when the text is difficult, we can expect gains at the textbase level of comprehension for the majority of readers (as in Experiment 2). If the text is overly challenging, of course, we would expect to see little to no effects of training (on that text).

Future Directions

Our future research will examine whether a greater amount of practice may allow the less skilled reader to achieve gains at deeper levels of comprehension. However, in order for this amount of practice to be available to the student, iSTART will need to be incorporated into the classroom. That is, the teacher will have to be able to use iSTART to fit the needs of the course curriculum. However, the current version of iSTART provides self-explanation practice with only two science texts. While this has been sufficient to conduct our initial evaluations of the system, it is not sufficient to take iSTART to scale.

Our research indicates that, the current version of iSTART must be augmented in two principle ways: (1) by improving the ability of the system to adaptively respond to students' and teachers' needs and, (2) by creating a teacher interface that allows teachers to integrate iSTART into their classrooms. Unfortunately, incorporating a computerized tutor into a classroom is not as simple as just giving it to the teacher and expecting that it be used consistently or successfully. There are many teacher needs that must be met to reach our goals. First, the teacher must understand the need for reading strategy training and be receptive to intelligent tutoring systems. Hence, we intend to seek funding to develop and

test an automated program which provides teachers with information about the importance of reading strategies and information about iSTART, particularly how it helps students to read and understand difficult material. The program will additionally facilitate the teacher's use of iSTART in the classroom. Second, teachers are pressed to cover content and many have dropped the deep-learning activities that take 'extra time.' Thus, the learning material covered in iSTART must be relevant to the course demands. Currently, iSTART tutors students using a limited number of texts, covering content that may or may not be relevant to the particular course. Therefore, we intend (with future funding) to increase the number of course topics that can be covered during demonstration and practice and increase the difficulty range of practice texts.

Increasing the number of texts is essential in order for struggling students to receive a greater amount of practice than currently possible. Although practice with two texts provides substantial benefits, our research has indicated that this is not enough practice for the students most in need of training. In addition, it has been evident, through our experiences with students and teachers over the past few years, that the reading strategy training offered by iSTART should be available within other domains such as history and literature. For example, these two topics have been consistently listed by over half (and up to 75%) of the high-school SERT participants in our studies as topics for which they have used SERT strategies (though they were trained with science texts).

Expanding the number of topics covered by iSTART texts will contribute in both applied and theoretical ways. First, teachers will be more receptive to using iSTART in their classroom if topic-relevant material (or better yet, selections from the course textbook) is available for the student to read on the iSTART system. We all know that teachers are pressed to cover a certain amount of content during the year, particularly given recent pressures toward meeting state standards. Thus, iSTART will be better received by teachers if it allows coverage of required content while their students are learning the reading strategies. In theoretical terms, increasing the number of topics covered will contribute to our understanding of reading processes because it will afford the opportunity to examine how different strategies are used and contribute to comprehension across different domains.

References

- Best, R. M., Ozuru, Y. & McNamara, D. S. (2003). Self-explaining science texts: Strategies, knowledge and reading skill. Paper submitted to the International Conference for the Learning Sciences, California, LA.
- McNamara, D. S. (in press). SERT: Self-Explanation Reading Training. *Discourse Processes*.
- McNamara, D. S., Levinstein, I. B. & Boonthum, C. (2003). iSTART: Interactive Strategy Trainer for Active Reading and Thinking. Submitted to *Behavioral Research Methods, Instruments, and Computers*.
- McNamara, D. S., Kintsch, E., Songer, N.B., & Kintsch, W. (1996). Are good texts always better? Text coherence, background knowledge, and levels of understanding in learning from text. *Cognition and Instruction, 14*, 1-43.
- O'Reilly, T., Sinclair, G. P., & McNamara, D. S. (2002). Reading Strategy Training: Automated versus Live. Proceedings paper submitted to The Sixth Annual International Conference of the Learning Sciences.