

FINDINGS

This section describes the principal results of our reading strategy training experiments conducted using SERT (experimenter-delivered training; McNamara, 2004) and iSTART (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 participant population, or particular experimental procedures. Findings from the teacher training and observation sessions are not included as we are currently in the process of collecting the data.

SERT: EXPERIMENTER-DELIVERED READING STRATEGY TRAINING

Assessing the Effects of SERT, Preview, and INSERT with High-School Students.

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). Last year we reported some of the principal findings from a subset of the experiments that were conducted (Chemistry, Honors Biology, and Earth Science). This year we completed the assembly a database for all three schools. This was a considerable undertaking as the final dataset was over 1.5 million data cells. An analysis across all three schools provides a more representative sample in order to test the effectiveness of the learning strategies.

Our primary question was to analyze the effects of training condition immediately after training by assessing students' comprehension on a passage about viruses. There were 1090 students (546 males, 544 females) who completed the virus comprehension test. A mixed analysis of variance (ANOVA) was conducted with question type (text-based, bridging) and question format (multiple-choice, open ended) as within-subject variables and the training condition as the between-subject variable. To control for differences in question format, the analysis was conducted using z-scores pertaining to performance on the virus questions. For multiple-choice questions, responses were scored as either correct or incorrect; however for open-ended questions, partial credit was given for answers that contained a portion of the correct answer. We also covaried the four individual difference measures (reading skill, domain knowledge, MSI, MARS) to determine whether there were any effects of condition over and above the impact of individual difference measures.

The analysis revealed a significant effect of condition, $F(3,960)=5.83$, $MSE=12.62$, $p<.05$. A post hoc least significant difference test revealed that SERT participants ($M=.175$, $SD=.73$) outperformed INSERT participants ($M=-.031$, $SD=.72$), Cohen's $d=.28$, and Preview participants ($M=-.072$, $SD=.73$), Cohen's $d=.34$; however SERT participants were not significantly different from control participants ($M=.113$, $SD=.73$). Moreover, control participants ($M=.113$, $SD=.73$) scored significantly higher than Preview participants ($M=-.072$, $SD=.73$), Cohen's $d=.25$.

There was also a significant interaction between condition and question format, $F(3,960)=6.72$, $MSE=4.03$, $p<.05$, indicating that SERT participants ($M=.27$, $SD=.98$)

scored significantly higher on open-ended questions than control participants ($M=.14$, $SD=.99$), $F(1,657)=2.80$, $MSE=2.73$, $p=.047$, Cohen's $d=.13$ (one-tailed test). In addition, there was a significant interaction between condition and question type, $F(3,960)=3.24$, $MSE=1.89$, $p<.05$, indicating that SERT participants ($M=.44$, $SD=1.75$) scored marginally higher on bridging questions than control participants ($M=.25$, $SD=1.67$), $F(1,662)=2.07$, $MSE=6.06$, $p=.075$, Cohen's $d=.11$ (one-tailed test). There were no other significant main effects or interactions.

In summary, with human delivered training, SERT is more effective in improving comprehension than INSERT or Preview training. Moreover, SERT is also more effective than no training control on more difficult open-ended questions, and marginally more effective on bridging-inference questions.

iSTART: INTERACTIVE STRAEGY TRAINING FOR ACTIVE READING AND THINKING

Assessing Changes in Reading Strategies as a Function of Live SERT and iSTART.

To compare the relative effectiveness of live (SERT) and computer-based (iSTART) reading strategy training, college students read and self-explained science texts prior to and after training. In Experiment 1, participants received SERT and in Experiment 2, participants received iSTART. Sentence-focused, local, and global self-explanations were assigned scores of 1, 2, and 3, respectively. In both Experiments 1 and 2, we calculated an average general self-explanation score for pre- and post-training sessions. Table 1 contains the data from both experiments. With respect to Experiment 1 (live SERT), a reading skill by test mixed ANOVA was conducted. There was a significant reading skill by test interaction, $F(1,26) = 8.12$, $Mse = .04253$, $p < .05$. Post-hoc analyses (LSD) revealed that skilled readers had higher strategy scores after training than before. In contrast, less-skilled readers had lower strategy scores than before training, although this difference was marginal ($p < .10$). Furthermore, prior to training, skilled and less-skilled readers had similar strategy scores, whereas after training skilled readers had higher strategy scores than less-skilled readers.

A reading skill by test mixed ANOVA was also conducted on the strategy scores in Experiment 2 with iSTART (see Table 1). There was a significant main effect of test such that the strategy scores were higher post-iSTART ($M = 1.92$) than pre-iSTART ($M = 1.80$), $F(1,48) = 3.86$ $Mse = .08150$, $p = .055$. A main effect of skill approached significance such that skilled readers ($M = 1.92$) had higher strategy scores than less-skilled readers ($M = 1.78$), $F(1,48) = 2.98$ $Mse = .205$, $p < .1$. The interaction between reading skill and iSTART was not significant.

Table 1. The mean strategy score as a function of reading skill and time.

STRATEGY	SKILL	TEST	
		Pre	Post
Experiment 1 (SERT)	Skilled	2.18 (0.38)	2.28 (0.32)
	Less-skilled	2.11 (0.52)	2.03 (0.40)
Experiment 2 (iSTART)	Skilled	1.85 (0.44)	2.02 (0.27)
	Less-skilled	1.71 (0.45)	1.81 (0.37)

NOTE: Standard deviations are reported in the parentheses.

Table 2 shows the LSA cosines for current sentence, prior text, and title in Experiment 1. A reading skill by source by time mixed ANOVA was conducted in which SERT and source were within-participants variables and reading skill was a between-participants variable. There was a main effect of test such that cosines were higher after training ($M = .43$) than before training ($M = .33$), $F(1,26) = 16.97$ $Mse = .02075$ $p < .05$. There was also a main effect of source, $F(2,26) = 116.86$ $Mse = .001066$, $p < .05$. Post-hoc analyses (LSD) revealed that cosines for the current sentence ($M = .47$) were higher than the prior text ($M = .35$) or title ($M = .32$), which did not differ.

Table 2. Mean LSA cosines as a function of skill, source, and SERT.

SKILL	SOURCE	SERT	
		Pre	Post
Skilled	CS	0.40 (0.15)	0.48 (0.15)
	PT	0.30 (0.13)	0.37 (0.11)
	Title	0.29 (0.15)	0.37 (0.12)
Less-skilled	CS	0.42 (0.13)	0.56 (0.12)
	PT	0.30 (0.07)	0.41 (0.07)
	Title	0.25 (0.12)	0.39 (0.08)

NOTE: Standard deviations are reported in the parentheses.

Table 3 shows the LSA cosines for current sentence, prior text, and title in Experiment 2 with iSTART. A reading skill by test by source mixed ANOVA was conducted. There was a main effect of test such that cosines were higher after training ($M = .42$) than before training ($M = .35$), $F(1,48) = 21.64$ $Mse = .01339$, $p < .05$. There was also a main effect of source, $F(2,48) = 161.22$ $Mse = .001809$, $p < .05$. Post-hoc analyses (LSD) revealed that cosines for the current sentence ($M = .50$) were higher than the prior text ($M = .33$) or title ($M = .33$), which did not differ.

Table 3. Mean LSA cosines as a function of skill, source, and iSTART.

SKILL	SOURCE	iSTART	
		Pre	Post
Skilled	CS	0.47 (0.16)	0.54 (0.13)
	PT	0.31 (0.09)	0.36 (0.06)
	Title	0.29 (0.10)	0.36 (0.06)
Less-skilled	CS	0.46 (0.18)	0.54 (0.15)
	PT	0.30 (0.10)	0.35 (0.08)
	Title	0.30 (0.09)	0.35 (0.09)

NOTE: Standard deviations are reported in the parentheses

The goal of this study was to compare the effectiveness of iSTART to live SERT for skilled and less skilled readers. We were particularly interested in changes in the general reading strategies and relevance of self-explanations as a function of training. With respect to live training, skilled readers engaged in more global processing after training than before, whereas skilled readers did not. However, both skilled and less-skilled readers produced self-explanations that more relevant to the local and global

contexts after training. What is encouraging about the current study is that we found that both skilled and less-skilled readers increased their strategy scores and produced more relevant self-explanations after iSTART than before. Taken together, these results suggests that both skilled and less-skilled readers change the way they read as a function of both live and computer-based strategy training.

Assessing Differences in Reading Strategies as a Function of Reading Skill and the Length of a Self Explanation.

To explore better ways to tailor feedback to students, this analysis was conducted to examine the quality of self-explanations as a function of the number of clauses in the explanation and the student’s reading skill. Our analyses are still ongoing and the results reported here are preliminary. Table 4 contains the mean strategy scores as a function of self-explanation length. A reading skill by length mixed ANOVA was conducted on the strategy scores, with reading skill as a between-participants variable and clause length as a within-participants variable. There was a significant effect of clause length ($F(1,41) = 41.257, Mse = .068, p < .05$). Post hoc tests (LSD) revealed that self-explanations with three or more clauses had higher scores ($M = 2.47$) than those with two clauses ($M = 2.17$) which in turn had higher scores than those with one clause ($M = 2.00$). This main effect was qualified by a significant reading skill by length interaction ($F(1,41) = 41.257, Mse = .068, p < .05$). Posts hoc test revealed that high skilled readers had higher strategy scores for self-explanations containing one and two clauses. However, there were no differences between readers of differing skills for self-explanation containing three or more clauses.

These results have implications for delivering feedback to readers of varying skills when using iSTART. Specifically, they suggest that less-skilled readers should be encouraged to produce longer self-explanations. Conversely, given that high skilled readers are able to generate more active reading strategies in linguistically compact self-explanations, they may be frustrated by blanket instructions to say more. These results ultimately suggest that iSTART would benefit from pre-training assessments of students. These assessments may be used to tailor feedback during training and practice.

Table 4. Mean LSA cosines as a function of skill and the number of clauses.

SKILL	NUMBER OF CLAUSES		
	1 Clause	2 Clauses	3 or > Clauses
Skilled	2.13 (0.35)	2.28 (0.27)	2.48 (0.30)
Less Skilled	1.86 (0.37)	2.05 (0.30)	2.46 (0.43)

NOTE: Standard deviations are reported in the parentheses.

Evaluation of iSTART v1.0 with Middle-School Students.

This year we analyzed the data collected from two versions of iSTART: version 1.0 and version 2.0. The results of the study with middle school students (iSTART v1.0) are presented first. The results of the comprehension component of this study were presented in last year’s report. However this year, we present the analyses performed on the quality of self-explanations with iSTART v1.0. Next we describe the results from an experiment conducted with high-school students who were trained on the new and

improved iSTART version 2.0. Finally we discuss the set of experiments for which we are currently collecting data utilizing iSTART v2.0. These experiments examine the effects of integrating iSTART into the classroom as well as the effects of repeated practice.

Our 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. We do not describe the full detail of the study in this section because a fully detailed description of the study was provided in the last year's progress report. Instead, we will report the results of a new analysis that emerged in the ongoing data analyses following the preparation of the last years' progress report.

Self-explanation Protocol Analyses

In the last year's progress report we reported the results of protocol analyses performed on thunderstorm text, one of the two texts used in the practice phase of iSTART training. We have completed protocol analyses on two more texts: 1) the coal and thunderstorm texts used in the practice phase; and 2) the heart disease text, a text used to measure the effect of training on the students' ability to self-explain without the presence of feedback. Thus, we separately describe the results of data analysis for: 1) combined data for the both coal and the thunderstorm texts; and 2) the heart disease text.

1. Analysis of the protocol produced for practice texts

We have scored (using the procedure described in the activities section of the 2003 progress report) 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, 2004). We focus here on the effects of individual differences and sentence difficulty on paraphrasing and elaboration, and the relations between paraphrase and elaborations. In general, students often used paraphrasing (in 91% of the self-explanations) and these paraphrases were often successful ($M = 0.94$). Paraphrases that used fewer words from the sentence (called distant paraphrases) occurred more frequently for high-knowledge than low-knowledge students (0.63 vs. 0.44), $F(1, 39) = 7.5$, $p < .01$.

Elaborations were less frequent than paraphrases, occurring in 41% of the self-explanations. More elaborations were produced by skilled than less skilled readers (0.49 vs. 0.30), $F(1,33) = 6.1$, $p < .05$. And, skilled readers produced more accurate elaborations than less skilled readers (0.59 vs. 0.49), $F(1,39) = 9.2$, $p < .01$. Of all elaborations produced, 70% were based on general knowledge/common sense, 21% were based on prior text, and 9% were based on scientific knowledge. The majority of the elaborations (i.e., 72%) focused on the current sentence, whereas only 10% focused on a global level of comprehension beyond the current sentence (knowledge building). Skilled readers' advantage in the production of elaborations over less-skilled readers was observed not only in sentence focusing elaboration (0.35 vs. 0.20), $F(1, 39) = 9.9$, $MSE = 0.043$, $p < .01$, but also in knowledge building elaborations (0.06 vs. 0.01), $F(1, 39) = 910.07$, MSE

= 0.006, $p < .001$. In sum, skilled readers produced higher quality elaborations, which contained more knowledge building, sentence linking, and were more accurate.

In addition, the analysis on the relations between paraphrase and elaborations indicated that high quality elaborations tend to co-occur with paraphrases; that is, the data indicated that 1) elaborations tend to be irrelevant when produced in isolation as opposed to with paraphrases (0.33 vs. 0.12), $F(1, 32) = 13.9$, $MSE = 0.056$, $p < .01$; and 2) a larger proportion of knowledge building and sentence focusing relevant elaborations, as opposed to irrelevant elaborations, occurred with paraphrases (0.91 vs. 0.73 vs. 0.50), $F(2, 24) = 7.3$, $MSE = 0.0757$, $p < .01$. Overall the data indicate that paraphrasing plays important role in bolstering the generation of quality elaborations.

2. Analysis of the protocol produced for test text (heart disease text)

The goal of this analysis was to assess the effects of iSTART training on students' ability to produce quality self-explanations. A series of Univariate ANOVAs were conducted with the proportion of paraphrases or elaborations (out of all the sentences self-explained) as the dependent variable and the training condition (iSTART, control) as the independent variable. There was an effect of training condition on the production of elaborations, such that students in the iSTART condition ($M = 0.36$, $SD = 0.25$) produced more elaborations than students in control condition ($M = 0.15$, $SD = 0.13$), $F(1, 37) = 11.75$, $p < 0.01$. Further, training affected the quality of elaborations produced, such that iSTART students produced more relevant elaborations ($M = 0.31$, $SD = 0.24$) than control students ($M = 0.12$, $SD = 0.12$), $F(1, 37) = 9.69$, $p < 0.01$.

In summary, students trained with iSTART were more likely to produce elaborations than control students, and the elaborations were of better quality. This is an important finding because elaboration activities are associated with subsequent comprehension success (McNamara, 2004; Chi, 2000). Our data suggest that students at the middle-school level benefit from self-explanation reading training, thus indicating the importance of providing students with self-explanation reading strategy training.

We also analyzed the relationships between self-explanation strategy use and comprehension success. We ran a series of correlations to explore associations between the production of elaborations and paraphrases and comprehension. There was a positive correlation between the production of elaborations and comprehension success, $r = 0.48$, $p < 0.01$, but the correlation between production of paraphrases and comprehension test was not significant, $r = 0.13$, $p = 0.40$.

A subsequent analysis of elaborations showed that the quality of elaborations are associated with comprehension; relevant elaborations were associated with general comprehension success, $r = 0.52$, $p < 0.01$, but irrelevant elaborations were not, $r = 0.01$, $p = ns$. Furthermore, relevant elaborations were associated with successful performance on the text-based questions, $r = 0.34$, $p < 0.05$, and the bridging inference questions, $r = 0.45$, $p < 0.01$. A closer examination of the data indicated that of the relevant elaborations, current sentence focusing elaborations were associated with the comprehension of bridging-inference questions, $r = 0.41$, $p < 0.01$, but not text-based questions, $r = 0.21$, $p = 0.18$. Conversely, knowledge building elaborations, which build a more global representation of the text were associated with comprehension of both the

text-based questions, $r = 0.46$, $p < 0.01$, and bridging-inference questions, $r = 0.34$, $p < 0.05$. In sum, our results indicate an important link between the production of elaborations and comprehension success among middle-school students.

Evaluation of iSTART v2.0 with High-School Students.

Our first experiment with version iSTART v2.0 investigated the efficiency with which the modified iSTART v2.0 program teaches high-school students to use the SERT strategies. In contrast to our previous investigations of iSTART, this study was conducted in the students' classrooms during real class time. One outcome of 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 were 405 Shelby County Tennessee high-school students. The classes, taught by 5 teachers, included 10 biology classes (10th grade) and 10 physical science classes (9th grade). All of the students completed the pretest aptitudes assessments in September, 2003, and were administered a posttest in April, 2004. Half of the classes were trained in November, 2003, and the remaining half in January, 2004. Using matched-samples assignment (based on reading skill and prior knowledge), half of the students in each classroom were assigned to the iSTART condition and half were assigned to a control condition.

The students in the control condition were 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 kind of control condition accounts for effects of novelty pertaining to computer use as well as exposure to scientific information involved in iSTART training. Following the iSTART and web design training phase, both the experimental group and control group completed 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 5. The two main dependent variables included immediate post-test and delayed post-test comprehension. The immediate comprehension test was on a passage on viruses and it was administered on the last day of training. The delayed post-test passage was on the origin of the universe and it was administered near the end of the academic year.

Table 5: 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

The Impact of iSTART in a Classroom Setting.

Our first question was to examine the impact of condition immediately after training. We performed a mixed ANOVA with question type (text-based, bridging) and question format (multiple-choice, open ended) as within-subject variables and condition (control, iSTART) as the between-subject factor. The analysis did not reveal any significant effects for condition $F(1,403) < 1$, nor were there any significant interactions with condition. The effects of condition were also not reliable when the individual difference measures (reading skill, MARSIs, MSI and domain knowledge) were entered as covariates $F(1,334) < 1$.

We performed a second mixed ANOVA on the delayed comprehension test at the end of the academic year including question type as the within-subject variable and condition as the between-subject variable and. We covaried out the four individual difference measures (reading skill, domain, knowledge, MSI, MARSIs) to determine whether there were any effects of condition over and above the impact of individual difference measures. There was a marginally significant effect of condition, $F(1,306)=3.62$, $MSE=.273$, $p=.058$, Cohen's $d=.16$ (significant at $p=.029$ using a one-tailed test), indicating that the iSTART participants ($M=.49$, $SD=.25$) scored higher than the control participants ($M=.45$, $SD=.25$) on the delayed comprehension test.

In summary, we did not find any significant effect of condition immediately after the training session. However we did find a significant effect of condition at the delayed post-test; iSTART participants scored significantly higher than controls on the comprehension questions pertaining to the universe passage. There may be several reasons why we observed an effect of training at the delayed post-test, but not after the immediate post-test. One possibility is that the virus text, used at the immediate post-test, was more difficult than the universe text, used at the delayed post-test. This hypothesis is supported by the fact that the average scores on the immediate text-based questions were lower ($M=.46$ $SD=.24$) than the scores on the delayed text-based questions ($M=.58$, $SD=.27$). Another possibility is that performance may be lower immediately after training because students are fatigued after the training. Finally, it may be that students practiced using the SERT strategies and thus became more proficient in using the sub-strategies during reading. The possibility of repeated practice is important; one

concurrent goal of our current research is to investigate the effects of student practice in comprehension.

Effects of Repeated Practice.

Now that we have obtained evidence to show that SERT and iSTART benefit students' ability to comprehend science texts, we are focusing on whether students' require extended training to become "better readers". We are in the process of conducting three experiments, henceforth called Experiments A, B, and C, to examine whether a greater amount of practice self-explaining science texts allows students to become better self-explainers and increases comprehension of science texts.

In order for regular 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. Hence, one activity was to increase the number of texts available for practice on iSTART.

Experiment A

Experiment A assesses the impact of weekly iSTART practice in which students practice self-explaining one science text per week from their regular course textbooks. Students are given one class period (around 50 minutes) per week to self-explain science texts which are incorporated to the iSTART Practice module. Each week, a second class period is used to assess students' comprehension of the science text they self-explained and the remainder of the chapter from which the self-explained excerpts were derived. Students completed a weekly multiple-choice quiz designed to assess textbook knowledge (see weekly quiz information in the Activities Section).

The participants include approximately 200 Shelby County Tennessee high-school students. The classes, taught by two teachers, include five life sciences classes (9th grade) and five physical science classes (9th grade) We are working with the Life Science classes in the fall semester (August-December, 2004) and the Physical Science classes in the spring semester (January – April, 2005). The experiment follows the procedure of pre-test (2 consecutive class periods), training (3 consecutive class periods), repeated practice (one class period per week for six consecutive weeks) and post-test (2 consecutive class periods).

Experiment A also examines the role of the iSTART system's feedback to assess the helpfulness of feedback in self-explanation quality. We therefore manipulated the feedback condition, such that students either received feedback or did not receive feedback. A between-subjects design is used in the Life Sciences classes; students either always receive feedback, or do not receive feedback in each practice session. Conversely, in the Physical Science classes, a within-subjects design will be used, such that students will receive feedback for some self-explanations, but not others.

Experiment B

Experiment B further investigates the importance of repeated practice on students' ability to self-explain and comprehend science texts. In this experiment we compare computer-delivered practice to teacher-delivered practice to examine whether differences in self-explanation/comprehension success emerge as a function of the "mode" of practice. The teacher training provides an opportunity to assess whether teachers can readily administer training.

The participants include 93 students Shelby County Tennessee high-school students. The classes, taught by one teacher include four Biology classes (9th grade). We are currently running Experiment B, which will be completed by December, 2004.

Experiment B is identical to Experiment A in all respects except the practice. Two of the Biology classes complete the weekly practice using iSTART with feedback whereas the remaining two classes experience classroom practice training in which the teacher leads the practice session using feedback. The teacher takes the students through the texts incorporated into the weekly practice iSTART module. Thus, the teacher asks students to produce self-explanations for sentences contained in the science text and provides feedback.

Experiment C

Experiment C examines whether self-explanation training can help the less skilled readers to achieve gains at deeper levels of comprehension. We investigated the benefits of the iSTART intervention and repeated practice training among students classified by their school as having low-level "reading comprehension skills". Specifically, we compared the quality of self-explanations and level of comprehension among students taking a special reading class (reading training students) and main-stream/honors students before and after the reading training students are trained with iSTART. Our hypothesis is that the comprehension gap between reading training students and main-stream students would reduce after the reading training students have completed the extended practice training. The main-stream students are not exposed to iSTART or repeated practice training.

The participants include 97 ninth-grade Shelby County Tennessee high-school students. The reading class comprises 15 students, which is taught by 1 teacher. The remaining students are regular/honors Life Science students taught by 1 teacher. We are currently running Experiment C which will be completed by November, 2004.

As the reading training students are "struggling" comprehenders we expect that they would benefit from regular, intensive practice. Following training with iSTART, students will undergo self-explanation practice training for one class period per day for three consecutive weeks using the plan outlined in Table 6. On day 1 of each week, students practiced self-explaining a text using the iSTART practice module; on day 2 the teacher provided classroom practice training using the same text; on the third day, students practiced self-explaining a second text using the iSTART practice module; on the fourth day the teacher provided classroom practice training for the second text. Finally, on the fifth day, the students completed a comprehension quiz which assessed their knowledge of the texts used in the self-explanation practice.

The kinds of text used in the practice weeks varied, such that students self-explained life science, literature and wellness texts in weeks 1, 2 and 3, respectively. We used texts from different content areas to examine whether the self-explanation strategies we teach students with science texts help students' better comprehend what they read in other subject domains. We believe that the investigation of self-explanation practice and subsequent comprehension in different subject domains is particularly important for "struggling readers" who are likely to encounter comprehension difficulties in all content areas.

Table 6: Weekly practice schedule

Day	Activity
Monday	iSTART practice (text 1)
Tuesday	Teacher practice (text 1)
Wednesday	iSTART practice (text 2)
Thursday	Teacher practice (text 2)
Friday	Weekly quiz

FUTURE DIRECTIONS

The experiments we have conducted and are conducting with SERT and iSTART are geared to providing us with a better understanding of how to improve students' ability to understand science texts. The work that we have conducted over the past year has also provided us with a better understanding of the changes we need to make to iSTART to ready it for the classroom, teacher, and student. We have a much better understanding have the difficulties we will encounter on a large scale, what teachers need in order to use the system, and how we can better tailor training to the individual. Our work over the next year will be to complete the current experiments in the classroom and to continue modifications of iSTART according to the guidelines yielded from our experimental findings. We intend to apply for a one-year no-cost extension for the 2005/2006 year. During that final year, we will complete our analyses, write-up our results for publication, and continue modifications of iSTART.