ACTIVITIES

INTRODUCTION

The ability to understand and learn from text is a skill required to function effectively in modern society. Unfortunately, however, many students have difficulty understanding what they read; in particular, many students have trouble comprehending science texts (Bowen, 1999; Snow, 2002). At least part of the problem is due to students' seldom using high-level comprehension strategies in the classroom (Cox, 1997; Garner, 1990). Indeed, one way to improve comprehension is to teach reading strategies that encourage deeper processing of the text (e.g., Chi, De Leeuw, Chiu, & LaVancher, 1994; McNamara, de Vega, & O'Reilly, in press; Pressley, Wood, Woloshyn, Martin, King, & Menke, 1992). The purpose of this project, "Promoting Active Reading Strategies," is to develop reading strategy training programs that can be incorporated into science classes to help students understand science texts.

The "Promoting Active Reading Strategies" project has been a five-year endeavor, designed to help high-school and college students better understand science texts by teaching them to use active reading comprehension strategies. One component of the project concentrated on Self–Explanation Reading Training (SERT) delivered by teachers. SERT requires readers' to self-explain text (e.g., Chi et al., 1994) using various reading strategies known to facilitate comprehension (e.g., logic and common sense, elaboration, and bridging inferences). Empirical studies (reported in the 2003 and 2004 Promoting Reading Strategies progress reports) indicate that SERT helps high-school and college level students better understand science texts (e.g., McNamara, 2004; McNamara & Scott, 1999). Furthermore, SERT training has been found to be particularly useful for facilitating comprehension among low-knowledge readers (O'Reilly, Best, & McNamara, 2004); low-knowledge readers trained to use SERT outperform control participants on measures of text comprehension, and perform as well as high-knowledge participants.

The second and most important focus of the project has been developing the automated version of SERT called iSTART (Interactive Strategy Trainer for Active Reading and Thinking; McNamara, Levinstein, & Boonthum, 2004). As reported in the 2003 and 2004 progress reports, we have been successful in developing iSTART (v1.0; v2.0). iSTART (v1.0 and v2.0) comprises 3 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 self-explanations while the program provides feedback on performance). We have also been successful in evaluating the efficacy of iSTART in laboratory and high-school classroom settings. As reported in the 2003 report, a laboratory experiment showed that iSTART (v1.0) was as successful as SERT in terms of facilitating the comprehension of scientific material among college students (O'Reilly, Sinclair, & McNamara, 2004). For high-school students, iSTART (v.1.0 and v2.0) was found to improve the quality of self-explanations that students produce for science texts in laboratory and classroom situations (Best, O'Reilly, & McNamara, in press; Todaro, Magliano, Millis, Kurby, & McNamara, 2004), which are associated with comprehension success (O'Reilly, Sinclair, & McNamara, 2004).

There were three goals in the past year's funding period: 1) experimentally evaluating the reading strategy interventions, 2) preparing to integrate iSTART into high school classrooms, and 3) improving the performance and the usability of the iSTART system.

The first goal was to experimentally evaluate the iSTART intervention. To this end, we conducted further tests of the effectiveness of iSTART by examining data from previous experiments. Specifically, we compared the quality of self-explanations produced before and after iSTART training to examine the effect of iSTART on the quality of self-explanations that students produce (Experiment 1). We also explored the relationship between self-explanation quality and comprehension question answering performance (Experiment 2), in particular, reading strategy use and reading comprehension.

Although self-explanation practice with two texts provides some benefits, such a small amount of training may not be sufficient for students to acquire reading strategy skills. Thus, we also pursued our first goal by assessing the contribution of extended practice in self-explanation quality and comprehension among the high-school population. We conducted four experiments (Experiments 3, 4, 5 and 6) in which students participated in extended practice sessions (e.g., daily sessions for three-week duration or weekly practice sessions for six-week duration), and compared how different types of extended practice influenced the students' mastery of reading strategies. To this end, we varied the extended practice conditions in several ways, including comparing computer practice to live practice (conducted by the teacher), comparing computer practice with feedback (normal practice) to computer practice without feedback, and comparing all-sentence to target-sentence practice. The latter experiment examined the effectiveness of having students self-explain selected "target sentences" of a science text (sentences deemed to be most critical to comprehension) as opposed to self-explaining every sentence of a text (as in the Practice module of iSTART training).

Our work on extended practice helped us advance toward our second goal of integrating iSTART (v2.0) into high-school classrooms by ensuring that iSTART can be administered by teachers in high school classes and used by students of differing ability levels, including "struggling comprehenders." In terms of teacher administration, the experiments with extended practice provided important data from teachers who led students through iSTART in the four experiments. We collected feedback from the teachers with regard to their opinions of and their likelihood of using iSTART. To examine generalization across student abilities, we conducted experiments with honors students, regular students, and students classified as experiencing "comprehension difficulties."

The third goal was to address several aspects of iSTART to improve the usability and performance of the system. First, efforts have been made to provide more flexibility to researchers (and teachers) for selecting the desired measures from a battery of testing items, and imposing constraints, if desired, on those measures (e.g., controlling test time or requiring completion of the current test before proceeding to the next test). Second, the evaluation procedures designed to classify the self-explanation protocols produced by students have been explored to: 1) determine if the evaluations can extend to texts other than the trainings texts, and whether more fine grained evaluation algorithms can improve the feedback system of the iSTART Practice module (and the post-training, extended practice); 2) assess whether the self-explanations are affected by reading skill and text characteristics (Experiment 7), and response mode effects on self-explanations (Experiment 8a and Experiment 8b). Third, we have improved the method of data extraction (e.g., self-explanation protocols, pre-and post-test data) to ensure that researchers can easily obtain specific data without altering the primary database.

WORK CONDUCTED

This section describes the work we have undertaken to meet the three goals which were specified in the previous section. Three aspects of the project central to the goal of this research are: (I) experimental tests of reading strategy interventions, including experiments concerning the factors influencing the effectiveness of self-explanation reading strategy intervention; (II) integrating iSTART into the classroom, and (III) improving the performance and usability of iSTART.

I. Experimental Tests of Reading Strategy Interventions

We have conducted a series of experiments to evaluate the effect of iSTART on students' ability to self-explain and comprehend science texts among samples of middle-school, high school, and college students (Goals 1 and 2). At the most general level, our experiments can be categorized as having a goal of either (1) evaluating the effectiveness of iSTART, or (2) evaluating the effects of extended practice. Although the two types of experiments serve slightly different purposes, they are structurally similar. Thus, we describe them together, and note differences when appropriate. Six experiments in this report were conducted to assess the effectiveness of reading strategy training based on iSTART: Experiment 1 examines the effectiveness of iSTART training on comprehension of science materials through improvement in students' ability to self-explain the science texts; Experiment 2 is a large scale high-school classroom study that assess the effect of iSTART on students' ability to self-explain science texts; Experiments 3, 4, and 5 assess the effect of different types of extended practice on students' ability to self-explain and comprehend science texts; Experiment 6 examines the effect of iSTART with extended post-training practice on reading comprehension ability among students classified as having "comprehension difficulties."

A. Overall structure of the experiments

All of our experiments follow the same structure, comprising pretest, reading strategy training, and posttest phases. During the pretest, students complete a battery of aptitude and achievement tests and a demographics questionnaire to establish whether the effects of reading strategy training varied across individuals with different aptitudes (e.g., prior knowledge, general reading comprehension skill, self-explanation ability). During the iSTART training phase, students are provided with instruction on how to use active reading strategies. Student comprehension of science texts and quality of self-explanations for science texts is assessed immediately after the training to investigate the impact of reading strategy training on the comprehension of science texts. Some experiments further incorporate an extended self-explanation practice in which students practice self-explaining texts on a weekly or daily basis (for a total of three to six weeks) after the iSTART training. At the end of the semester or academic year, students complete a posttest consisting of a battery of aptitude and achievement assessments, most of which were administered during pre-test. In addition, teachers answered a questionnaire that assesses teacher demographics and teaching and testing behaviors. Finally we have also collected course performance measures such as grades; however, some of this data is not yet available for this report.

B. Extended practice

Experiments 3, 4, 5 and 6 assessed the effects of extended, post-training practice on self-explanation quality and science text comprehension (see Findings Section). We varied the extended practice in three ways to assess how different forms of extended practice affect the students' learning of the reading strategies: 1) we compared computer-based practice to teacher-guided practice (Experiment 3); 2) we compared conditions in which feedback on performance is provided during practice (as in the iSTART Practice module) to a no feed back condition (Experiment 4); 3) we compared practicing self-explanation on selected sentences of a text to self-explaining all sentences of a text (Experiment 5); and 4) we observed the effects of a combination of teacher-guided and computer-based extended practice for students classified as having "comprehension difficulties." Extended practice was conducted on a daily (Experiment 6) or weekly (Experiments 3 through 5) basis (see Findings Section for details).

C. Control groups

Experiment 1: A reading control group was used in which the students read and were tested on the same texts to which iSTART-trained students were exposed.

Experiment 2: Students completed a web-design task in which they designed web pages using the same texts that iSTART students were exposed (see below for details)

Experiments 3 through 5: Control groups were not used. Experiments 1 and 2 established that iSTART impacts reading comprehension skill by improving the students' use of reading strategies. Thus, the focus of these experiments was on extended practice beyond that of the iSTART training itself, and the conditions under which this extended practice optimizes self-explanation quality (e.g., practice with feedback compared to practice with no feedback).

Experiment 6: The progress of the iSTART-trained students of Experiment 6 will be compared to non-trained students from the general student population with respect to reading comprehension ability.

D. Experimental materials

iSTART training was provided via computer using either iSTART v1.0 (Experiment 1) or iSTART v2.0 (Experiments 2, and 3 through 6). The data from pre-test, training, and post-test were collected using conventional paper and pencil tests (some pre- and post-test components of Experiments 1 and 2) and with computerized tests, using a recently developed automated testing system (Experiments 1 through 6; please see previous reports that describe this system).

E. Pre- and Post-training measures

(1) *Demographic information*

A 26-item demographics questionnaire is 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.

(2) *Prior Knowledge*

A multiple-choice prior knowledge test is used to assess general knowledge relating to the disciplines of biology, chemistry, research methods, earth science and math. The test consisted of 20 items.

(3) *Science vocabulary knowledge*

We developed a 20–item multiple-choice test to assess students' comprehension of scientific vocabulary. This test was developed by selecting science terms from the students' science textbook.

(4) *Metacognition and reading strategy knowledge*

We have used two measures of students' knowledge of reading strategies. The majority of our experiments have used the Metacomprehension Strategy Index (MSI; Schmitt, 1990). The MSI is a 25-item multiple-choice questionnaire designed to measure various components associated with reading strategy knowledge such as predicting and verifying; previewing; purpose setting; self-questioning; drawing from background knowledge; and summarizing. In 2003, we developed a shortened, more robust version of the scale that includes 9 items (see 2003 progress report, Section III). Furthermore, in the iSTART experiments we have used the Meta-cognitive Awareness of Reading Strategies Inventory: (MARSI): a 30-item questionnaire requiring students to indicate the frequency with which they use particular meta-cognitive strategies using a five-point Likert scale (i.e., never to always).

(5) *Reading comprehension skill*

General reading skill was measured using a modified version of the standardized Gates-MacGinitie reading skill test for grades 10-12. The test consisted of 40 multiple-choice questions designed to assess student comprehension on several short text passages.

(6) *Science comprehension ability*

Students read a science passage and answered comprehension questions about the passage. For example, one such text/test for high-school students describes the nature of earthquakes. The passage is 336 words in length with a Flesch Reading Ease of 61.19 and a Flesch-Kincaid Grade Level of 8.2. Comprehension is assessed using open-ended questions. One half of the questions are text-based (the answer can be derived from a single sentence in the text) and the other half require bridging inferences (information from two or more sentences must be combined to arrive at the correct answer).

Comprehension texts and tests are presented on computer. The program first displays the text in its entirety. Once a student has read the text, they proceed to answer the questions. Due to class-time constraints, the comprehension tests are time-limited (15 minutes to read the text and answer the questions). Hence, the program is designed to allow the student to go back and review the text while answering the test questions. Both the review feature and the time-limit feature can be turned off. Additionally, the time allotted is selected by the researcher.

(7) *Self-explanation ability*

In Experiments 2 through 6, we administered a self-explanation test in which students produced self-explanations for sentences contained in short science texts. The tests were administered at pre-test, immediately after the training, and at post-test.

For Experiment 2, students provided a self-explanation for each sentence of a six-sentence text. The self-explanations were handwritten. For Experiments 3 through 6, students self-explained eight "target sentences" from a 20-sentence 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 the text to the students 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.

F. Assessments of the effects of the training

(1) *Science comprehension*

Immediately following training, students' comprehension of science passages was assessed in the manner described in section E (6) above. The particular text used varied as a function of target population and experiment.

(2) *Self-explanation*

Quality of student self-explanations was assessed in the manner described in Section E (7) above. Students were required to write (Experiment 2) or type (Experiments 3 through 6) self-explanations for sentences of a brief science passage.

(3) *iSTART Introduction module quizzes*

Student answers to the strategy quizzes completed during the Introduction module of iSTART training were collected to assess learning during the training period.

(4) *Post-training iSTART Quiz*

Shortly after iSTART training (one day to one week depending on the experiment), a 12-question, multiple-choice quiz was administered to assess students' memory for the strategies learned during training.

G. Assessments of the extended practice training (Experiments 3 through 6)

(1) *Self-explanation*

Quality of self-explanations was assessed in the following manner. The teachers were asked to select one section of text (around 20 sentences) from their science textbook for each week of extended practice. The students typed self-explanations for the sentences or target sentences (depending on the experiment) of the selected passage during extended practice (with the exception of the teacher-guided practice condition of Experiment 4 in which self-explanations were given verbally and not recorded). Thus, students practiced self-explaining portions of text from a chapter they were actually covering in class during the weeks of the extended practice sessions.

(2) *Extended Practice comprehension*

For each teacher-selected section of text, a six to eight multiple-choice quiz was constructed. The quizzes were administered by the teacher one to four days following the completion of practice for that section (i.e., when convenient for the teacher).

(3) *Science chapter knowledge*

To further investigate the effects of extended practice on comprehension, students' knowledge of the general chapter content, was assessed. Scores from chapter tests (produced and administered by the teachers) were obtained for each chapter from which an extended practice text section was extracted.

H. Experiments 3, 4, 5 and 6: Effects of Extended Practice

In the past, we have obtained evidence that SERT and iSTART improve the ability to comprehend science texts through improvement in reading strategy use. Currently, we are examining whether different types of extended self-explanation practice further enhance comprehension and learning from texts. This year, we conducted four experiments, referred to here as Experiments 3, 4, 5, and 6, to examine how different types of extended practice help high-school students become better self-explainers and increase their comprehension of science texts. As indicated in last year's progress report, in order for regular practice to be available to the student, iSTART needs to be integrated into the classroom. This means that science teachers need to be able to manage the training and subsequent practice sessions. Thus, teachers conducted the training and the extended practice sessions in Experiments 3, 4, 5 and 6. Note that teachers were trained to administer iSTART (and subsequent practice sessions). They were also exposed to cognitive psychological theories of reading comprehension during the June 2004 summer workshop (see Activities Section) to prepare for the training and practice sessions. For the extended practice component, teachers selected texts from their regular science textbooks, which they normally covered in the course curriculum.

The goals of these experiments were twofold. First, we assessed how different conditions of the extended practice (e.g., computer versus live practice) supplement the basic training. In other words, we identified optimal methods of extended practice that buttress the iSTART training for facilitating self-explanation skills and comprehension. The second goal was to collect data on how science teachers handle the training themselves. The experiments were conducted with ninth grade science students from a Shelby County school. The followed procedure was used: (a) pre-test (two consecutive class periods), (b) training, post-training test (four consecutive class periods), (c) extended practice (Experiments 3, 4, 5 - one class period per week for six consecutive weeks; Experiment 6 - four class periods for three consecutive weeks),and (d) post-test (two consecutive class periods).

For the extended-practice sessions, students were given one class period (approximately 50 minutes) per week to self-explain science texts that were incorporated into the iSTART Practice module, or as a class (teacher-guided conditions of Experiments 3 and 6). Each week, a second class period was used to assess comprehension of the text they had self-explained. Students completed a weekly multiple-choice quiz designed to assess textbook knowledge (see weekly quiz information in the Activities Section). Below, we outline the specific manipulations relating to the extended practice sessions for each experiment.

The data for all experiments is currently being processed and is not available for this report.

(1) *Experiment 3: Computer-based versus teacher-guided extended practice*

In this experiment we compared computer-delivered practice to teacher-guided practice in four biology classes to examine whether there would be a differential effect of extended practice type. The extended-practice sessions also provided an opportunity to compare the teacher's ability to instruct students under both practice conditions.

The participants were 90 Shelby County Tennessee high-school students drawn from four biology classes (all taught by the same teacher). All students participated in the computer-based strategy training (iSTART v2.0). Two classes participated in computer-based extended practice (presented in the same manner as the iSTART Practice module, with feedback). Two classes participated in teacher-guided practice sessions. In the teacher-guided sessions, the teacher asked students to self-explain the texts, calling on several students to provide self-explanations for each sentence of the text. The students self-explained orally, rather than typing self-explanations into the computer on an individual basis. The teacher was given few instructions by the iSTART researchers for the practice sessions, other than to use the specified texts, to go over the texts on a sentence-by-sentence basis, and to encourage the use of all of the iSTART reading strategies. The teachers were also asked to offer feedback to orally presented self-explanations; the methods and details of feedback was determined by the teacher. By allowing teachers to organize practice sessions, we can evaluate the efficacy with which teachers conduct the practice sessions and identify the types of teaching activities (e.g., discussions) and tasks that work well in group learning contexts.

(2) *Experiment 4: Extended practice with feedback versus extended practice without feedback*

Experiment 4 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 in which it was presented in the iSTART v2.0 Practice module during training, or they received no feedback during the extended practice phase (using the practice module of iSTART v2.0 with the feedback mechanism turned off).

The participants were 164 Shelby County Tennessee high-school students in five life science classes (all taught by the same teacher). Three of the classes received computer feedback whereas two classes did not receive computer feedback.

(3) *Experiment 5: Target sentence versus all sentence extended practice*

Experiment 5 compared a target-sentence practice condition to an all-sentence practice condition during the extended practice phase. In this experiment, for each practice text, sentences (approximately one-half of the total number of sentences per text) deemed most critical to understanding a text were pre-selected as "target sentences." Students in the target-sentence practice condition read the entire text (using the Practice module of iSTART v2.0, with feedback) but were instructed to provide self-explanations for the target sentences only when they appeared (identified as targets with text highlighting). Students in the all-sentence condition read and provided self-explanations for all sentences in the texts (using the practice module of iSTART v2.0, with feedback).

The participants were 144 Shelby County Tennessee high-school students in five physical science classes (taught by the same teacher). Two of the classes practiced self-explaining in the target-sentence condition, two of the classes practiced self-explaining in the all-sentence condition, and one class was assigned to a control condition. Control students participated in the pre- and post-testing phases of the experiment, but did not participate in iSTART v 2.0 training or the weekly practice sessions; rather, they participated in regular classroom teaching activities and were not told about the iSTART reading strategies.

(4) *Experiment 6: Benefits of iSTART training and extended practice for low-level comprehenders*

The final experiment assessed whether self-explanation training with iSTART can help high-school students identified as having reading comprehension problems gain the reading comprehension skills necessary to understand high-school texts. Specifically, we compared the quality of self-explanations and level of comprehension for high-school texts between students taking a special reading class (iSTART-trained students) and regular students (mainstream/honors students who did not receive iSTART training).

Moreover, we expanded the domains of the practice texts to include literature and health and wellness (part of the physical education curriculum) in addition to science.

The participants were 97 ninth-grade Shelby County Tennessee high-school students. Fifteen of the students were enrolled in a special reading class based on their reading comprehension scores from an eighth grade exit exam. These 15 students participated in all phases of the experiment. The remaining 82 students participated in the pre- and posttest phases of the experiment. They also read and were tested for comprehension on the texts used for practice in the iSTART-trained condition. They did not participate in any strategy training or extended practice sessions.

As the iSTART-trained students are "struggling" comprehenders we expect that they would benefit from intensive regular classroom based practice. Following training with iSTART, students' undergo self-explanation practice training for one class period per day for three consecutive weeks as outlined in Table 1. For each week, two texts were selected to be explained from the domain being covered that week (science, literature, or health and wellness).

Table 1: Weekly Extended Practice Schedule for Experiment 6

Monday: Computer-based practice on Text 1 using iSTART v2.0 Practice Module
Tuesday: Teacher-guided practice on Text 1
Wednesday: Computer-based practice on Text 2 using iSTART v2.0 Practice Module
Thursday: Teacher-guided practice on Text 2
Friday: Weekly quiz on Texts 1 and 2.

On Day 1 of each week, students practiced self-explaining a text using the iSTART Practice module with feedback. On Day 2 the teacher guided a classroom practice session using the same text. On Day 3, students practiced self-explaining a second text using the iSTART practice module with feedback. On Day 4 the teacher guided a classroom practice session for the second text. Finally, on Day 5 the students completed a comprehension quiz which assessed their comprehension/memory for the two texts.

The topic domains of texts used during extended practice varied. 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 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.

Experiment 6 required the reading class teacher to run the training and practice sessions. Again, the teacher was trained to administer iSTART (and subsequent practice sessions) and given training on theories of reading comprehension during the June 2004 summer workshop (see below).

II. Integrating iSTART into the classroom

Integration of iSTART into to high-school classroom involves interfacing with teachers who actually lead and control the training process in the classroom.

A. Interfacing with teachers

To integrate iSTART into the high-school classroom, it is necessary to train teachers to administer the training, and to examine how our system can be adapted to the teachers', students', and schools' needs and constraints. This year, we have worked toward integration by setting up teacher training prior to the in-class testing and collecting opinions on the iSTART training (see outline in progress report 2004) following the in-class testing.

B. Teacher training

During the summer vacation period (June 2004), we held a three-day training session with eight teachers from the two Shelby County high schools participating in our experimental research. The purpose of the training was to educate teachers about iSTART and the way in which the program may be used in the classroom. On Day 1, the teachers were given the theoretical background to the development of iSTART (including the review of models of text processing). On Day 2, the teachers worked through the three modules of iSTART (Introduction, Demonstration and Practice). On Day 3, the teachers completed open-ended questions in which they were asked to indicate: 1) the extent to which the training helped them understand theories of text comprehension and the theoretical stance leading to the development of iSTART; and 2) their opinions of iSTART as a classroom teaching tool. The structure of the training session is outlined in last year's progress report.

C. Teacher-delivered training

Given that teachers will be leading the training in the classroom, it is important that we evaluate the ways in which teachers use the system. We need to identify problems (e.g., technical, content-related problems with iSTART, or student motivation problems) that teachers and students encounter during teacher-guided training in the classroom. Further, we need to establish whether computerized training is sufficient, or whether it is better augmented by classroom training sessions. Thus, we are currently investigating the ways in which teachers deliver the training and weekly practice sessions, and the extent to which students benefit from computer-assisted practice and teacher-guided practice. We are working with four of the eight teachers who attended the summer training course. The four teachers guided students through iSTART and the extended practice training for Experiments 3 through 6 (see findings section) using the training procedures outlined above.

(1) *Training materials*

(a) *Training script*. Teachers were provided with a training script that provided instructions on how to guide students through iSTART. The script offered information about the purpose of iSTART, logging-on instructions, and procedures for seeking help should students encounter difficulties.

(b) *Teacher questionnaire*. Teachers completed a short checklist-style questionnaire after each class training period which assessed the types of problems students encountered and the frequency of each problem. Problems were classified as technical (e.g., logging on), procedural (e.g., following iSTART instructions), content-related (e.g., understanding meanings of words in the program), and other iSTART-related problems pertaining to the Introduction, Demonstration and Practice modules. Teachers were also asked to indicate whether they encountered problems taking students through iSTART, and their perceptions of students' motivation to use the program.

(c) *Observer questionnaire*. Observers from the University of Memphis completed a short checklist-style questionnaire after each class training period which assessed the types of problems students encountered and the frequency of each problem. Observers also evaluated each teaching session in terms of problems encountered and students' motivation and enjoyment.

(2) Practice materials

(a) *Practice texts*. A University of Memphis representative met with the teachers to discuss the format of the weekly practice sessions. As mentioned in the experimental section, extended practice was conducted on the computer (iSTART Practice module format), or in the form of classroom teaching sessions. For the computerized practice, teachers went through the log-on procedure with the students and answered questions relating to the practice. For the teaching sessions, teachers were asked to conduct a "whole class" training session in which students practiced self-explaining the texts as a group. Each student was given a copy of the text to be self-explained. The teacher's role was to elicit self-explanations from members of the class and provide appropriate feedback.

(b) *Teacher questionnaire*. Teachers completed a short checklist-style questionnaire after each practice period to evaluate the session in terms of problems encountered and student motivation.

(c) *Observer questionnaire*. Observers from the University of Memphis completed a short checklist-style questionnaire after each practice session which assessed 1) the classroom teaching format and 2) the types problems teachers encountered and the frequency of each problem. For the former, the focus was on the style of teaching (e.g., discussions, computer activities and class activities). For the latter, the focus was on the types of information teachers provided (e.g., information about the strategies used in iSTART, content pertaining to the texts being self-explained) when problems were encountered.

III. Improving the Performance and Usability of iSTART.

Regarding Goal 3, the researchers at Old Dominion University and Northern Illinois University have focused their efforts on improving the performance and usability of iSTART in three ways: 1) Improving the ease of configuring experiments (Experiment setup tool); 2) improving the manner in which the student self-explanations are evaluated; and 3) improving the methods of data extraction. For the first goal, efforts have been made to provide more flexibility for researchers to selectively administer, and impose constraints on, a battery of testing assessments. The second aim has been geared towards improving the automated evaluation of student self-explanations. Progress has been made towards developing more precise evaluation algorithms to improve feedback and to assess the appropriateness of the algorithms to evaluate the pre- and post-test self-explanation texts. The third aim is directed at facilitating the ease with which data is extracted. The improvement in data extraction has occurred at all levels including self-explanation protocols, post-training iSTART quizzes, individual difference measures, and comprehension tests.

A. Experiment Configuration and Experiment Setup Tool

The setup of experiments has been only partially automated. The pre- and post-tests are configured by an administrative interface, but the overall schedule of modules has been maintained manually in a database table. We are in the process of developing a tool that allows a researcher to configure an experiment by choosing the sequence of modules for each group (condition) in the experiment. The tool allows for the copying of one group's schedule to another and the setting of scheduling rules (*e.g.*, may/must/may not complete a module after scheduled date.) This tool will form a basis in a reduced form for part of a future teacher interface that will allow the teacher to select the modules and sequence of modules desired when using iSTART in the classroom.

B. Self-Explanation Protocol Evaluation

(1) *Revisiting the Evaluation Algorithm*

We currently have two sets of criteria by which humans evaluate self-explanation protocols. The original is a single dimension 4-point scale that evaluates the overall quality of the protocol with integer values ranging from 0 to 3. More recently, we have employed a multi-dimensional coding scheme that provides a more fine-grained analysis. Our current automated evaluation algorithm successfully predicts the protocol quality based on the 4-point scale. One goal is to develop an algorithm that successfully predicts more detailed qualities of the self-explanations as reflected in the multi-dimensional coding scheme. For example, the multi-dimensional coding scheme explicitly codes whether there was paraphrasing or elaboration, and the type of elaboration in the protocol. Our initial analysis indicated that the current algorithm based on word-based matches and LSA-benchmark cosines did not successfully predict the multi-dimensional coding results. We continue our efforts to develop an algorithm that can identify more fine-grained aspects of self-explanation protocols. Nonetheless, in the course of this exploration, we encountered two issues with respect to the LSA space: missing words and misspellings.

(a) *Missing words in the LSA space*. The formulas for evaluating the LSA space are based on an exact match of the words involved. When a new text is added to our library of practice texts, there is no guarantee that all of the words will be found in the space. *A fortiori*, there is no guarantee that every word in the student protocols will be found in the space. When words are missing in the space, the evaluations simply ignore their existence in the text or protocol and therefore yield somewhat impoverished results. Our initial plan to deal with this problem was to substitute an appropriate word that does exist in the space for a missing word when possible. The appropriate word would be discovered by adding or subtracting suffixes such as 's', or 'ed', or 'ing'. We searched the space to discover cases where such pairs already existed in order to see how appropriate such substitutions might be. When we compared the first ten dimensions of the LSA vectors for such pairs we found that there is no consistent relation between them: frequently corresponding values in the vectors have different signs or values of significantly differing magnitude. Rather than proceed with our original plan, we are now considering a second solution that involves building an LSA space from the same corpus after reducing each term in the corpus to its lemma (base form) and evaluating the texts and protocols after they have been similarly reduced. Discussions with colleagues lead us to believe that such a matrix should have more dimensions than are found in the current space. Unfortunately, determination of the appropriate number of dimensions requires a corpus of tuning data that we have not yet been able to discover. We are also planning experiments to examine the degree to which the missing words detract from the predictive accuracy of our method.

(b) *Misspelled words*. This is related to the previous issue since a misspelled word is one that is missing from the LSA space. To catch a misspelling that matches an existing entry would require sophisticated parsing that would add significantly to the program's response time. We found a spell-checker (*Aspell*) that seems to provide an appropriate list of possible corrections but we need to test it before deciding to incorporate such spell checking. Do we simply take the first spelling that matches a word in the LSA term-list or the one with the smallest Levinshtein distance to the misspelling? Does the addition of a spell-checker actually improve results? Is the performance penalty too high? These questions will be addressed as we continue to develop our evaluation algorithm.

(2) *Preliminary Evaluation of Self-Explanation Tests*

At three points during the curriculum (before the training, immediately after training, and after the series of weekly practices) the students were given two texts to self-explain with no coaching feedback. These protocols need to be evaluated using our algorithms to determine whether the quality of their self-explanations improves by those measures. The texts are less than half the length of those used in the practice and weekly practice modules so we are not certain that our formulas can simply be transferred. At this point we have used the current practice evaluation algorithm for the SE test protocols from Fall 2003. If this evaluation is not valid, we will use a different evaluation algorithm. Otherwise, we can use this evaluation for the Fall 2004 and Spring 2005 data.

(3) *Exploring Methods of Using LSA*

We are exploring the utility of using LSA in different ways to help classify self-explanations. Typically, the LSA cosines between the self-explanations and semantic benchmarks use all dimensions in the LSA space. It has been noted in a recent conference on LSA that the values on some dimensions are redundant, and when included in the computation of cosines, can lead to increased error. For example, if a dimension has the same value in the vectors representing each of two benchmarks, then the cosine between a self-explanation and each of these two benchmarks might more similar than if that dimension was not used. We tested the idea that correlations between cosines and human judgments would increase if similar dimensions are omitted. Specifically, we rank-ordered the dimensions on standard deviations computed from the elements that corresponded to words being compared. A low standard deviation would indicate similar values on the dimensions, indicating redundancy whereas a high standard deviation would indicate different values and uniqueness. We then omitted 0%, 15%, 20%, 25% and 30% of the dimensions that had the lowest standard deviations, and computed cosines between two segments of text (self-explanations and benchmarks) using the remaining dimensions. The correlations between those cosines and human judgments of similarity were: .31, .29, .26, .24, and .19, respectively. Therefore, instead of increasing the power to detect differences between protocols, omitting the dimensions only diminished it.

(4) *Reading Skill, Text, and Response Mode Effects on Self-Explanations*

(a) *Experiment 7 Text constraints and reading ability*. One of the goals for creating a flexible feedback system is to ensure that the algorithm used to assess self-explanation quality can offer appropriate feedback for a wide range of users. We are examining the appropriateness of feedback used in iSTART by evaluating the self explanations of both skilled and less skilled readers, including information contained in a particular sentence, the student's prior knowledge, or information contained in different sentences. This information will be useful for iSTART in that if a student supplies an explanation lacking expected information, then its feedback might include or be influenced by that observation.

In this study, skilled and less-skilled college students, as defined by performance on the Nelson-Denny test of comprehension, supplied verbal protocols after each sentence of two expository texts. The content nouns for each protocol were classified based on the type of information. If the noun was included in the current sentence (the sentence that had elicited the protocol), then it was classified as current sentence. If the noun had been mentioned earlier in the text, then it was classified as prior text. If the noun did not appear in the text at all, then it was classified as world knowledge. Generally, current sentence is associated with paraphrasing, prior text with bridging, and world knowledge with elaborating. For each sentence, the following five predictor variables were determined: number of new argument nouns, prior knowledge, number of local and distal causal connections, and number of nouns in the sentence that had appeared earlier in the text (argument overlap). New-argument nouns were defined as the number of nouns in the sentence that had not appeared before in the text. Prior knowledge scores assessed the extent that the reader had prior knowledge relevant to the current sentence. The scores were cosines generated by LSA which compared the student's prior knowledge of the topic (ascertained by a prior knowledge test) and the sentence. Local causal relationships were present when a given text sentence was causally related to its immediate prior sentence. Distal causal relationships were present when a given text sentence was causally related to sentences that had occurred two or more sentences back in the text.

(b) *Experiment 8a and 8b Response mode: Talking or typing.* Traditionally, a significant amount of research on self explanations involves having participants produce their self-explanations by talking aloud. However, students using iSTART type their self-explanations into the computer. It is important to consider whether these two methods produce similar or different outcomes. It may be the case that typing thoughts requires more working-memory resources than talking aloud, leading to fewer cognitive resources allocated to using active reading strategies. Typing thoughts then would lead to worse self-explanations than talking aloud, underestimating their potential. Conversely, it could be the case that writing thoughts requires more self-editing and more carefully crafted self-explanations than thinking aloud. Writing then would lead to better self-explanations than talking aloud. Although this outcome could over-estimate their use of reading strategies, it does have the added benefit of the student producing and being recognized by iSTART for good self-explanations. Of course, it is possible that the two methods have the same impact on producing self-explanations. It would be beneficial to the goals of iSTART if there were (a) no differences between the modalities, or (b) an advantage for writing one's thoughts.

Two experiments have been conducted with college students. In Experiment 8a, participants wrote and spoke their self-explanations while reading expository texts, and in Experiment 8b, participants did so while reading narrative texts. Participants were also administered the Nelson-Denny reading skills test because potential differences between the two procedures might be mediated by the reading skill of the student. We are comparing reading strategies used by skilled and less skilled readers for each modality. Specifically, we are currently conducting a detailed analysis of the self-explanation protocols. We are also assessing various measures of comprehension. For the expository experiment (Experiment 8a), comprehension was measured both with free recall and short answer questions. For the narrative experiment (Experiment 8b), comprehension was measured by having participants recall the passages in as much detail as possible.

C. Data Extraction

(1) *Researcher Protocol Evaluation tool – Extracting Protocols*

We recognized a problem in the collection and evaluation of protocols. Student self-explanation protocols are evaluated in order to improve our algorithms and to assess the success of the training. The protocols are collected online in a database, but researchers prefer evaluating them in a spreadsheet as a preliminary to analyzing their results in SPSS. The problem is that the evaluations are not returned to the database. Other researchers may not know they exist or where to find them. In order to make the evaluations available to all researchers, we are in the process of building a tool that will allow the researcher to download data from the database in the form of an excel spreadsheet to which columns have been added for the purposes of the evaluation. The data in the spreadsheet is locked to prevent alteration. Once the researcher has entered evaluations of the protocols, the spreadsheet can be uploaded to the iSTART website where the evaluations will be incorporated with the original protocols. The features include: 1) The spreadsheet can be created on a per-student or per-question basis; 2) The evaluator can use any number of columns for evaluation; 3) Several evaluators can upload evaluations—all will be retained in different columns; 4) Evaluators must specify the meanings of the columns.

(2) *Extraction of individual differences and performance data*

A variety of reports have been generated for the pretest/posttest data, reading comprehension test, module quizzes, and practice data. In each case, data that had been collected in the database was manipulated and transferred to a spreadsheet. These transformations are accomplished by a set of Java programs that we are in the process of making more general and user-friendly.

(3) *Data Analysis – Extracting overall student performance*

We are in the process of extracting the data over the entire curriculum for each student in the Fall 2004 Millington experiments (Experiments 4 and 5). We have currently collecting the Introduction Quiz scores, the average level attained by the student in Demonstration, and quality of self-explanation protocol scores from the Practice module texts. However, we still need to incorporate the pretest and posttest data. Once collected, we will be able to measure changes in student test scores that might be attributed to the training and also determine whether any of the earlier data is predictive of later data. In the latter case, we may be able to use the data initially collected as a student profile to use in adapting the training to the student.

References

Best, R. M., O'Reilly, T. P., & McNamara, D. S. (in press). Comprehending science texts: Effects of computerized reading strategy training. *11th Biennial European Association for Research on Learning and Instruction Conference*, Nicosia, Cyprus.

Best, R., Ozuru, Y., & McNamara, D. S. (2004). Self-explaining science Texts: Strategies, knowledge, and reading skill. In Y. B. Yasmin, W. A., Sandoval, N. Enyedy, A. S. Nixon, F. Herrera (Eds.), *Proceedings of the sixth international conference of the Learning Sciences: Embracing Diversity in the Learning Sciences* (pp. 89-96). Mahwah, NJ: Erlbaum.

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

Cox, B. (1997). The rediscovery of the active learner in adaptive contexts: A developmental-historical analysis of transfer of training. *Educational Psychologist, 32*, 41-55.

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

Levinstein, I. B., McNamara, D. S., Boonthum, C., Pillarisetti, S. P., & Yadavalli, K. (2003). Web-based intervention for higher-order reading skills. *Proceedings of the ED-MEDIA 2003 World Conference on Educational Multimedia, Hypermedia &Telecommunications* (pp. 835-841).

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

McNamara, D. S., de Vega, M., & O'Reilly, T. (in press). Comprehension skill, inference making, and the role of knowledge. In F. Schmalhofer & C.A. Perfetti (Eds.), *Higher level language processes in the brain: Inference and comprehension processes*. Mahwah, NJ: Erlbaum.

McNamara, D. S., Levinstein, I. B. & Boonthum, C. (2004). iSTART: Interactive Strategy Trainer for Active Reading and Thinking. Submitted to *Behavioral Research Methods, Instruments, and Computers, 36*, 222-233.

O'Reilly, T., Best, R., & McNamara, D. S. (2004). Self-Explanation Reading Training: Effects for Low-Knowledge Readers. In K. Forbus, D. Gentner, T. Regier (Eds.), *Proceedings of the 26th Annual Meeting of the Cognitive Science Society* (pp. 1053-1058). Mahwah , NJ : Erlbaum.

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

O'Reilly, T., Sinclair, G. P., & McNamara, D. S. (2004). iSTART: A web- based reading strategy intervention that improves students' science comprehension. Proceedings of IADIS International Conference: Cognition and Exploratory Learning in Digital Age.

Ozuru, Y., Best, R., & McNamara, D. S. (2004). Contribution of reading skill to learning from expository texts. In K. Forbus, D. Gentner, T. Regier (Eds.), *Proceedings of the 26th Annual Meeting of the Cognitive Science Society* (pp. 1071-1076). Mahwah, NJ : Erlbaum.

Pressley, M., Wood, E., Woloshyn, V. E., Martin, V., King, A., & Menke, D. (1992). Encouraging mindful use of prior knowledge: Attempting to construct explanatory answers facilitates learning. *Educational Psychologist, 27* (1), 91-109.

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

Todaro, S. A., Magliano, J. P., Millis, K. K., Kurby, C. C. & McNamara, D. S. (2004). Intra-clause constraints in think-aloud protocols. In K. Forbus, D. Gentner, T. Regier (Eds.), *Proceedings of the 26th Annual Meeting of the Cognitive Science Society* (pp. 1642). Mahwah, NJ: Erlbaum.