

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).

One way to improve comprehension is to teach reading strategies that encourage deeper processing of the text (McNamara, de Vega, & O'Reilly, in press). Such interventions have been successful (e.g., Chi, De Leeuw, Chiu, & LaVancher, 1994; Pressley, Wood, Woloshyn, Martin, King, & Menke, 1992). In particular, McNamara (2004) has reported learning gains from Self-Explanation Reading Training (SERT). During SERT, students learn how to self-explain text (e.g., Chi et al., 1994). They also learn to use various reading strategies, particularly logic and common sense, elaboration, and bridging inferences, to improve their ability to explain text and understand it at a deeper level. McNamara's research (McNamara, 2004; O'Reilly, Best, & McNamara, 2004; O'Reilly, Sinclair, & McNamara, in press) has found that low-knowledge readers trained to use SERT outperform control participants on measures of text comprehension, and perform as well as high-knowledge participants.

There were two specific goals in this past year's funding period. First we wanted to further determine whether the experimenter-delivered training was effective for high-school students. To this end, we have continued to assess the effect of SERT on students' reading comprehension ability by comparing effects of SERT to other reading comprehension interventions and a control condition. A large-scale SERT study was conducted in three high-schools; a school with suburban and primarily white upper middle class student population; an inner city school with primarily African American and lower middle class students; and a rural Appalachian school with primarily white students. Our goal this year was to further examine this data by compiling a data set that includes all three schools.

Our second goal was to evaluate the effectiveness of automated SERT training using the Interactive Strategy Training for Active Reading and Thinking (iSTART; Levinstein, McNamara, Boonthum, Pillarisetti, & Yadavalli, 2003; McNamara, Levinstein, & Boonthum, 2004). iSTART, which is designed to be used in the classroom, is an interactive program; animated agents respond to the student by providing an appropriate level of feedback and scaffolding. As reported in the last year's progress report, we developed iSTART v2.0 by modifying the original iSTART, v1.0. We intend to expand iSTART v2.0 to provide a wide range of reading strategies to prepare for the integration of the system into the high-school classroom. As a step toward this goal, this year's activities include: 1) an analysis of the effect of iSTART v1.0 on the middle-school students' ability to self-explain science texts; 2) an assessment of the effect of iSTART v2.0 on the high-school students' science text comprehension; 3) conducting a more extensive training intervention studies with iSTART v2.0, which involve extended weekly reading strategy training in high-school classrooms, and 4) examining the impact of iSTART on improving the quality of student self-explanations, as well as exploring

different methods of maximizing the utility of LSA for assessing the appropriateness of the feedback.

Over the past two years we have assessed the effectiveness of iSTART v1.0 in both college (O'Reilly, Sinclair, & McNamara, 2004) and secondary school populations (O'Reilly, Sinclair, & McNamara, in press). Results from our research indicate that iSTART is as effective as SERT in terms of improving students' comprehension of science texts (O'Reilly, Sinclair, & McNamara, 2004). Across a number of experiments, we have found that iSTART and SERT improve comprehension at the textbase level for less strategic readers and at the situation model level for more strategic readers (O'Reilly, McNamara, & Sinclair, in press). Further, we established that iSTART teaches students to produce better quality self-explanations for science texts (Best, O'Reilly, & McNamara, submitted; Best, Ozuru, & McNamara, 2004; Todaro, Magliano, Millis, Kurby, & McNamara, 2004; Ozuru, Best, & McNamara, 2004) which are associated with comprehension success (O'Reilly, Sinclair, & McNamara, in press). The quality of self-explanations, however, is linked to student abilities, such that students with higher knowledge and reading skills produce "better" self-explanations (Ozuru, Best, & McNamara, 2004). Thus, the way in which students benefit from iSTART and SERT depends on their zone of proximal development.

In short, last year's progress report revealed that iSTART version v1.0 is effective in improving science passage comprehension at both the college level and secondary school level. Thus, our activities this year focus on the assessment of iSTART v2.0 in promoting comprehension and self-explanation quality. In particular, our focus was on assessing the effectiveness of iSTART in real-world classroom environments, as opposed to laboratory settings.

Finally, 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. Thus, the focus of much of this year's research is on the contribution of extended practice among the high-school population. Our work on extended practice has allowed us to work toward our ultimate goal of integrating iSTART (v2.0) into high-school classrooms.

WORK CONDUCTED

This section describes the work we have undertaken to meet our goals. Three aspects of the project central to the goal of this research are: (a) experimental tests of reading strategy interventions, (b) taking iSTART into the classroom (c) development and modifications of the iSTART system, and development of assessment techniques. Each is described in the three sections that follow.

I. Experimental Tests of Reading Strategy Interventions

We have conducted a series of experiments to evaluate the effect of various reading strategy interventions on the comprehension of science texts among samples of middle school, high school, and college students. At the most general level, our experiments can be categorized as: (1) the evaluation of the effectiveness of experimenter-delivered reading strategy interventions, and (2) the evaluation of computerized reading strategy training (iSTART).

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 depend on individual difference factors (e.g., prior knowledge and reading comprehension skill). During the training phase, students are provided with instruction on how to use active reading strategy (see below for details). Student comprehension of science texts is assessed immediately after training to investigate the impact of reading strategy training on the comprehension of science texts. Some of our experiments also incorporate refresher training in which students are reintroduced to the reading strategies shortly after the main training phase. In such cases, science text comprehension is again assessed immediately after the refresher training. Further, some of the experiments incorporate repeated self-explanation practice in which students practice self-explaining texts on a weekly or daily basis (for a total of three to six weeks). Finally, at the end of the semester, or academic year, students complete a posttest consisting of a battery of aptitude and achievement assessments. Teachers are also given a questionnaire that assesses teacher demographics and teaching and testing behaviors. We have also collected course performance measures such as grades; however, some of this data is not yet available for this report.

Reading strategy training

This section outlines the nature of the reading strategy training used in the experiments. These interventions are: 1) three types of experimenter-delivered reading strategy interventions (pre-view, SERT, and INSERT); and 2) iSTART, the computerized training.

Experimenter-delivered training

We investigated the effectiveness of three key types of theoretically grounded reading strategy training techniques, Self-Explanation Reading Training (SERT), Previewing, and the Interactive Notation System for Effective Reading and Thinking (INSERT). These interventions provide students with instruction to use comprehension strategies, and the training session takes approximately two hours.

(a) SERT

SERT is based on *self-explanation*, whereby students explain a text to themselves using a variety of reading and thinking strategies known to improve comprehension. These strategies are: (a) comprehension monitoring: being aware of one's understanding; (b) paraphrasing: restating the text in different words, and (c) elaboration: adding information to make sense of the text, such as using prior knowledge or previous information cited in the text. SERT helps comprehension by training students to first identify comprehension failure, and then repair the comprehension gaps by utilizing their prior knowledge and reasoning skills to repair any breaks. Overall, SERT is a high-level metacognitive reading strategy designed to improve comprehension by training the reader to make connections between the text and prior knowledge to promote deep comprehension.

SERT participants are first provided with a description and examples of self-explanation. The instructor then defines and provides examples for five reading strategies: comprehension monitoring, paraphrasing, elaboration/logic and common sense, prediction, and making bridging inferences.

Participants then watch a video depicting a student reading and self-explaining a text about forest fires. The video is paused at various points, and participants identify and discuss the strategies being used by the reader in the video. Next, the participants work in pairs to practice self-explanation while reading a science text (e.g., a chapter from their

science textbook). The participants take turns self-explaining, alternating after each paragraph. At the end of each paragraph, the partner who was listening (and not self-explaining) summarizes the paragraph.

(b) Previewing

Previewing is a reading strategy designed to improve comprehension by training students to preview various subsections of the text before they read. Students are asked to preview the title, introduction, pictures, figures, bold or italicized words, conclusion, and chapter review questions. Based on their preview, students are asked to write down what they know and what they need to know. Next, the students read the text and make notes regarding what they found out. Previewing is expected to improve comprehension by helping the student form a mental model of what they expect the text to be about.

Participants are first provided a description of the basic previewing strategy and subsections to preview (title, introduction, objectives bold italics, pictures/figures, conclusion and chapter review questions) and the notes sheets (What I know, What I need to know. What I found out). The instructor then demonstrates the strategy with a text on forest fires (as in SERT). Finally, the students practice the strategy with a science text (e.g., a chapter from the students' textbook).

(c) INSERT

INSERT involves recording symbols next to each paragraph to indicate agreement, importance, novelty, and understanding. The seven INSERT symbols include: '/' = I agree with what is being said; 'X' = I disagree with what is being said; '+' = this is new information; '!' = WOW, how interesting; '?' = I wonder; '??' = I do not understand; and '*' = this is important information. Participants are provided with a description and examples of the INSERT strategy along with a handout. The students practiced using this strategy with a science text.

INSERT is intended as a control condition in the sense that the students learn a new strategy, but it only involves learning to monitor comprehension and does not provide direction for repairing misunderstanding or provide help in mental model formation. While all three strategies training emphasize the importance of comprehension monitoring, they do so in different ways. Both SERT and INSERT promote online comprehension, while Previewing focuses on comprehension before the students read the texts. Only Previewing and SERT help the student build mental models of the material and provide guidance to fill in gaps in the model.

iSTART: Interactive Strategy Trainer for Active Reading and Thinking

iSTART is an interactive, automated reading strategy trainer that incorporates all the elements included in SERT (McNamara et al., 2004). iSTART consists of three major components: Introduction, Demonstration, and Practice. The first component introduces students to the five reading strategies: Comprehension Monitoring, Paraphrasing, Prediction, Elaboration and Bridging. Automated pedagogical agents (a teacher named Julie and two student assistants named Sheila and Mike) guide students through the introduction. Julie introduces and explains the five strategies one at a time, while Sheila and Mike provide examples of how each strategy can be used in generating a self-explanation. Students complete a short quiz after being introduced to each strategy to assess and promote their learning of the strategies. The quizzes are designed to guide the student to the correct understanding of each SERT strategy. The introduction is designed

to be self-paced in that students are given sufficient opportunity to read and understand the examples of self-explanation and to work through and learn from the quizzes.

In the second component, two new agents, Merlin and Genie, demonstrate the ways in which the five SERT strategies can be put to use when reading science texts. A science text is presented on the computer screen one sentence at a time. After reading a sentence aloud, Genie speaks aloud a self-explanation for the sentence. Genie's self-explanation then appears on the computer screen. After each explanation, students are asked to indicate the strategies that Genie used. Additionally, students are asked to identify the parts of Genie's self-explanation in which the indicated strategies occur and sometimes to find related parts of the science text. This procedure continues on a sentence-by-sentence basis.

The final component of iSTART allows students to practice self-explanation using the five SERT strategies. Students are presented with two science texts ('Stages of thunderstorm development' and 'The origin of coal') one sentence at a time. For each sentence, they are asked to type a self-explanation. The quality of the self-explanation is assessed by iSTART, and students are provided feedback by the agent Merlin based on this assessment. The feedback is designed to encourage students to continue to develop their self-explanation skills. For example, Merlin might respond with "Try thinking back to other sentences" if it is determined that a student tends to base his self-explanation on only the current sentence.

Repeated practice

Some of our iSTART assess the effects of repeated practice on self-explanation quality and science text comprehension (see Findings Section). Extended practice is completed on a daily or weekly basis (see Findings Section for details). To test the effectiveness of the system's feedback in eliciting "good quality" self-explanations and science text comprehension, we manipulated feedback, such that students either received feedback (as used in the practice module) or no feedback.

Finally, to establish the helpfulness of computerized practice in students' ability to comprehend science texts, we compared regular iSTART practice (using the practice module) to teacher-led practice (in which the classroom teacher takes students through practice). See Findings Section (Experiments A, B and C for details).

Control groups

Unless specified differently, we used a reading control condition in which students' read texts used in the training. Control students were generally told that the purpose of the study was to determine the types of strategies students use when they read.

Experimental materials

The measures used to assess student knowledge, at each phase of testing, are reported in this section. The assessments are paper and pencil tests and have been completed in booklets. However, we have recently developed an automated testing module (see Section III).

Pre- and Post-training measures

(a) 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.

(b) Prior Knowledge

A multiple-choice prior knowledge test is used to assess general knowledge relating to disciplines such as science, research methods and math. We developed two versions of this assessment – one for college students (comprising 54 items) and another for high-school students (comprising 35 items). For the iSTART experiments, we developed a shortened prior knowledge test, consisting of 20 items.

(c) Science vocabulary knowledge

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

(d) 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. We have recently developed a shortened, more robust version of the scale that includes 9 items (see last year's 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).

(e) Reading comprehension skill

Two standardized tests were used to assess reading comprehension skill. For college students, reading skill was measured with the Nelson Denny Reading Skills Test. The test consists of 38 multiple-choice questions designed to assess comprehension on several short text passages. For high-school students, reading skill measured with a modified version of the 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.

(e) Science comprehension ability

Students read science passages and answered comprehension questions about the passage. For example, one such test for high-school students consists of two texts, one about petroleum and another about the carbon cycle. The petroleum passage was 366 words in length, with a Flesch Reading Ease of 41.1 and Flesch-Kincaid level of 10.9, described how petroleum is refined. The carbon passage, comprising 320 words, with a Flesch Reading Ease of 51.0 and Flesch-Kincaid Grade level of 10.2, described the structure of plants and their role in the carbon cycle.

Comprehension is assessed via multiple-choice and/or open-ended questions. Half of the questions are text-based and the other half require bridging inferences.

(f) Self-explanation ability

In the iSTART experiments we administer a self-explanation test in which students write self-explanations to sentences contained in short science texts. This test is administered before and immediately after the training, and at the end of the school year.

Assessments of the effects of the training

(a) Science comprehension

Immediately following training, students' comprehension of science passages is assessed. The particular passage used depends on the target population. For a number of the studies with high-school students, we have used a text about viruses. The virus text describes the structure and reproduction of viruses as well as some examples of viruses and how they relate to disease. The passage is 1216 words in length with a Flesch Reading Ease of 45.1 and Flesch-Kincaid Grade level of 10.6. There is a set of 8 corresponding-open ended and 8 multiple-choice questions; half of which are text-based while the remaining half require bridging-inference. We also asked students to indicate whether they used any reading strategies while reading the science passages.

(b) Self-explanation

In several experiments, we have assessed the quality of student self-explanations. Students are required to write or type self-explanations to each sentence of a brief science passage.

Assessments of the extended practice training (Experiments A, B and C only)

(a) Self-explanation

We assess the quality of student self-explanations produced in the iSTART practice module. Students are required to type self-explanations to each sentence of a short textbook passage selected by the teacher. Teachers are asked to select a section of text (around 20 sentences) from their regular textbook. The aim is to select a portion of text from a chapter they are covering during the weeks of the extended practice sessions.

(b) Passage comprehension

We assess students' knowledge of the passages used in the self-explanation practice (computerized practice or classroom teaching practice). Eight multiple-choice items pertaining to each passage are administered within one to four days following the completion of the practice (i.e., when convenient for the teacher).

(c) Science chapter knowledge

To further investigate the effects of weekly practice training in students' comprehension success, students' knowledge of the general chapter content from which self-explanation excerpts were derived is assessed. We use scores obtained from the weekly quiz (produced by teachers) pertaining to the chapter from which the self-explanation passages were extracted.

Teacher Questionnaire

At the end of the school year, teachers are given a five-page survey that assesses their level of experience, familiarity with reading strategy interventions, and teaching techniques. The questions assessed items related to keeping their training updated (e.g., hours of continuing education, number of journals read); the nature of homework assignments (e.g., how much reading is assigned to students; hours of homework); use and knowledge of reading comprehension strategies; use of teaching aids, testing methods; and teaching attitudes.

Assessing the Quality of Self-Explanations Produced by Talking Aloud or Typing into the Computer

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 require 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 ones 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 also 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) least no differences between the modalities, or (b) an advantage for writing one's thoughts.

Two experiments have been conducted thus far on a college student population addressing the differences between written and spoken self-explanations. In Experiment 1, participants wrote and spoke their self-explanations while reading simple narrative texts, and in Experiment 2, participants did so while reading difficult scientific texts. Participants were also administered the Nelson-Denny reading skills test because potential differences between the two procedures may 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 conducting a detailed analysis of the self-explanation protocols, which is still ongoing. We are also assessing various measures of comprehension. For the narrative experiment (Experiment 1), comprehension was measured by having participants recall the passages in as much detail as possible. For the expository experiment (Experiment 2), comprehension was measured both with free recall and short answer questions. At this point in time, the analyses are ongoing and are not yet completed.

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

Self-Explanation Reading Training (SERT) has been shown to improve the comprehension of readers, and in particular less skilled or low-knowledge readers. A number of studies recently conducted have shown similar improvements in comprehension for students using iSTART. We have conducted a study specifically designed to assess changes in reading strategies as a function of live and computer-based training, which has been written up and submitted for publication (Magliano, Todaro, Millis, Wiemer-Hastings, Kim, & McNamara, under review). Improvements in reading comprehension demonstrated in prior research should be based on changes in reading strategies. We were particularly interested in how skilled and less skilled readers responded to both

The purpose of this study was to compare the relative effectiveness of live (SERT) and computer-based (iSTART) reading strategy training. Prior to and after training, college students read scientific texts and self-explained after each sentence. In Experiment 1, participants received SERT and in Experiment 2, participants received iSTART. Reading skill was assessed using the Nelson-Denny reading skill assessment tool. An important test of iSTART is to demonstrate that readers modify their strategies as a function of training, as is the case with live SERT. However, it is quite possible that we would find skilled and less-skilled readers practicing different strategies after training. Specifically, it is reasonable to expect that skilled readers would use global strategies more often than less-skilled readers because they should be more likely to emphasize the situation model (i.e., a "deep" representation about the situation model described by the

text). In contrast, less-skilled readers may practice paraphrasing and local processing strategies because these strategies help construct the textbase representation (i.e., a representation of the text). These expectations correspond to research regarding live SERT, which demonstrates that low-knowledge readers show benefits in a more accurate textbase post training, whereas high-knowledge readers show more accurate situation models after training. We would expect similar patterns with iSTART if it functions similarly to live training.

We assessed changes in comprehension in two ways. The first involved human judgments of general reading strategy. Our coding scheme has been applied in other studies conducted in conjunction with this grant. Two independent raters categorized the self-explanations as blank/vague, sentence-focused, local-processing, or global-processing. A blank/vague protocol either contained no response, was a semantically depleted utterance (e.g., “OK,” “uh huh,” “Yep”, etc), or an uninformative metacognitive statement (e.g., “I didn’t know that”, “I knew that,” etc). Sentence-focused explanations contained only clauses that were either partial or entire paraphrases of the current sentence. Local-processing explanations usually contained paraphrases as well, but included at least one word/clause that indicated either a “bridge” from the current sentence to the immediately prior sentence, or a minor elaboration based on world knowledge. Global processing explanations usually contained multiple clauses that indicated a bridge(s) to the distal text context (i.e., at least two sentences back from the current sentence) and/or major elaborations based on domain specific or domain relevant word knowledge. Furthermore, if a self-explanation contained more than one clause that reflected a SERT strategy other than a paraphrase (e.g., logic, bridge, elaboration, prediction), it was also judged to be global-processing. Inter-rater reliability in determining reading strategies was high (Kappa = .81). The judge was blind to whether the protocols were from a pre-SERT or post-SERT session.

We conducted another analysis to further explore how readers of differing skill responded to training. LSA was used to assess changes in strategies as a function of skill and training. We calculated cosines between the pre-SERT and post-SERT self-explanations and three information sources: current sentence, prior text context (i.e., prior two sentences), and the title. The prior text context reflected the local context, whereas the title was a proxy for the global context. That is, readers who were making local causal connections should explain how a sentence is related to the immediately prior sentence, whereas readers who were connecting the sentence to the global theme should mention information that is closely related the contents of the title. This approach has been successfully used to expose reading strategies.

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

One goal of iSTART is to ultimately be able to tailor feedback to the needs of the student. This agenda would be facilitated by an understanding of how readers of varying skill produce self-explanations. We have been analyzing pre-training self-explanations to further understand differences in the self-explanations produced by skilled and less-skilled readers. We were particularly interested in understanding properties of self-explanation of varying length. It may be the case that skilled readers are more able to produce short self-explanations that reflect more active reading strategies than less skilled readers. In order to address these issues, self-explanations have been identified in terms

of the number of clauses that they contain. Specifically, we identified clauses that had contained one clause, two clauses, or three or more clauses. We identified the global reading strategies reflected by the self-explanation using the coding scheme described above.

Exploring Methods of Using LSA

We are exploring different methods of maximizing the utility of LSA for estimating the semantic similarity between the self-explanations and the semantic benchmarks. Currently, the LSA cosines between the self-explanation and the semantic benchmarks are computed using all dimensions in the LSA space. Values on some dimensions of LSA are redundant, and can lead to increased error when included in the computation of cosines. For example, if a dimension has the same value in the vectors representing each of two semantic benchmarks, then the cosine between a self-explanation and each of these two semantic benchmarks might be more similar than if that dimension was not used. We are examining whether omitting such dimensions magnifies the differences between semantic benchmarks, thereby improving our classification of self-explanations.

This activity involves a systematic deletion of dimensions between two semantic benchmarks that fall below a threshold of similarity. Cosines are then computed between each semantic benchmark and a self-explanation. We then predict human ratings of similarity between the semantic benchmark and the self-explanation, as well as its human-given rating of general strategy. We are in the midst of this activity, and therefore, do not have any data to report at this time.

Optimization of the Weights Applied to Cosines Entering into the Choice of Feedback

LSA cosines between student thoughts and benchmarks are multiplied by weights that give some cosines more weight than others. These weights play a critical role in maximizing the reliability of the feedback system. The current set of weights was derived using discriminant analysis, and leads to acceptable feedback accuracy. Our goal was to improve the system's feedback accuracy by applying nonlinear functions to the cosines to derive better weights.

Neural networks can be used to obtain cosine weights through nonlinear transformation functions (e.g., sigmoidal, logarithmic, tangential). Weights obtained from this procedure may be better than linear weights for cosines. For example, a sigmoidal function increases variability of values in the mid range. Since most cosines lie in the midrange, this may increase their sensitivity to distinguish similar benchmark matches. Neural network training is a convenient method because the network develops the weights automatically.

A variety of transfer functions, network architectures with varying complexities, and other parameters were trained to identify weights that may offer a better match of the LSA cosine system with expert feedback.

None of the network configurations resulted in an improved classification over the current weight set. Instead, performance consistently hovered near the current classification performance, which is an acceptable value. This result shows on the one hand that the current weights are "as good as they get"; on the other hand the results show that the limitations to the LSA-based feedback selection procedure may not be easily rectified through weights.

Description of Specific Experiments

The findings section describes experiments assessing experimenter-delivered reading strategy training (SERT, Previewing, INSERT) and experiments assessing computer-delivered training using iSTART.

To summarize the results of experiments assessing experimenter-delivered and computer-delivered self-explanations, we have found that both human delivered SERT training and computer delivered iSTART training improve self-explanation quality and comprehension compared to participants who are not given training. More specifically, for the human delivered training, overall SERT participants out performed both Preview and INSERT participants on measures of immediate text comprehension. SERT participants also outperformed control participants on more difficult measures such as open-ended and bridging-inference questions. With regard to computer deliver training, our results indicate that iSTART was more effective in promoting comprehension at a year end delay than control participants. Moreover, in a separate experiment, iSTART participants were more likely to produce more effective self-explanations such as elaboration than control participants.

II. Integrating iSTART into the classroom

The influence of teacher delivered training; practice; and expanding the subject domain.

One of the key goals of the project is to integrate iSTART into the high-school classroom. To meet this goal it is necessary to train teachers to administer the training, and to examine how our system can be adopted to the teachers', students' and school's needs and constraints. This year, we have taken first steps toward integration by 1) training teachers to administer reading strategy training using iSTART and 2) making iSTART-related training accessible to a wider range of subject domains.

Teacher training

During the summer vacation period (June 2004) we held a three-day training session with 8 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. The structure of the training session is outlined in Table 3.

Table 3. Teacher training schedule

DAY	Activities
1	Theoretical background to the development of iSTART (models of text processing)
2	Working through iSTART modules (Introduction, Demonstration and Practice)
3	Evaluation session (Teachers discuss opinions about iSTART) Planning classroom training sessions for 2004-2005 academic year

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-led 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 or teacher-classroom practice. We are working with four of the eight teachers who attended the summer training course. The four teachers agreed to take their students through iSTART and the extended practice training for Experiments A, B and C (see findings section) using the training procedures outlined above.

Training materials

(a) Training script

Teachers were provided with a training script which provided instructions about how to take 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 kinds of problems students encountered (and the frequency with which each kind of problem was encountered). Problems were classified as technical (e.g., logging on), procedural (e.g., following iSTART instructions), content-related (e.g., understanding meanings of words in 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 kinds of problems encountered (and the frequency with which each kind of problem was encountered; see above). Observers also evaluated each teaching session in terms of problems encountered and students' motivation and enjoyment.

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), or in the form of classroom teaching sessions. For the computerized practice, teachers went through the logging on procedure with the students and answered questions relating to the practice training. 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) kinds of reading strategy information covered problems teachers encountered (and

the frequency with which each kind of problem. Section 1 monitored the style of teaching (e.g., discussions, computer activities and class activities). Section 2 monitored the kinds of information teachers provided (e.g., information about the strategies used in iSTART, or content pertaining to the texts being self-explained).

III. Changes to iSTART System and testing assessments

This section describes the changes made to the iSTART system, extensions of the automated testing assessments, and the development of practice self-explanation module that have been developed over the course of the past year.

1. Revision of evaluation algorithms.

A major goal of this phase of iSTART has been the reduction of human effort required to introduce new texts into the system for use in the practice phase. Much of our work has involved the revision of the algorithms used to provide a basis for feedback to the students. These algorithms evaluate the quality of the explanations the students produce and allow the system to respond with requests for revision, words of praise, or scaffolding remarks as is appropriate. In previous versions of iSTART we progressed from a word-based system to one which combined elements of a revised word-based system with elements of an evaluation based on Latent Semantic Analysis. In the latest version of the system, we revised the LSA approach to involve much less human effort.

- a. The original word-based system required a human expert to identify significant words and to develop an extensive dictionary for use by word-matching algorithms. Developing such a dictionary required approximate a week of intensive effort and had the drawback of requiring intuition as well as science.
- b. The system was revised to eliminate almost all human effort. A dictionary was created based entirely on the text in question and significant words were identified automatically. By itself this system performed less adequately than the original, but when combined with elements of the original LSA based system, the combined system outperformed both the LSA system by itself and the original word-based system by itself.
- c. The original LSA system required a human expert to perform a causal analysis on the text to generate a prior-text benchmark for each sentence and moderate scale gathering of protocols from subjects to develop a world knowledge benchmark for each sentence. Experimentation with revised formulas using our database of rated explanations helped us to discover a new and far less intensive approach. We replaced the original prior-text benchmark with one based simply on the prior two sentences and we eliminated the world knowledge benchmark entirely. After the system was recalibrated, it performed approximately as well as its predecessor. We also tried improving this system by adding words gathered from text using a semi-automated web search but found no improvement.
- d. In order to evaluate alternatives, a large set of protocols from previous experiments were evaluated by human experts for quality using the scale whose values are returned by the automatic evaluators. A variety of

statistical measures were used to compare the results of the automatic evaluator on the dataset with those of the human.

2. Authoring tool for new practice texts.

As stated above, the need to add many new texts has been an important goal of the past year. The human effort required to prepare the texts for evaluation has been reduced as described. In addition we have reduced the amount of work required to prepare the texts for presentation by developing an authoring tool. This tool is possible because the practice module has matured to the point where no modification of the code of the module is needed to present a new text. The text and all its additional annotations are stored in the central database. This authoring tool proceeds in several steps:

- a. The title and the entire text are pasted into the tool.
- b. The tool separates the text into sentences automatically and allows the user to correct the tool's mistakes (e.g., misinterpreting the period at the end of an abbreviation) or to combine several sentences into one item.
- c. The tool assists the author in identifying and correcting any mispronunciations made by the text to speech engine. Since the correction of mispronunciations is something of an art, the tool allows the author to identify the mispronunciations so that someone with more experience can make the actual correction.
- d. In addition to the pronounced version of the sentence, the tool will allow a presentation version of the sentence—one that includes HTML markup—to be created. (Under development.)
- e. The tool pre-computes the LSA vectors for the sentences of the text.
- f. Finally the several versions of each sentence (textual, pronunciation, presentation) are stored in the database.

3. Data conversion.

The iSTART system collects all data from tests and training in a MySQL database. The data are collected in a variety of formats appropriate to the particular module involved. For example, a student's set of answers to each multiple-choice pretest is preserved as a string of characters representing the student's answers, while each attempt at answering a multiple-choice question in the introduction section of the training is preserved as a separate record in the database. Researchers need to have the data for an experiment available in a form suitable for entry into a data analysis program such as SPSS. Consequently we have undertaken the effort to translate what is in the database in a form suitable for data collection into spreadsheet data suitable for analysis. Each major data format requires separate development.

- a. We have developed the means to provide the transformed data for the pre- and post-tests. The data is provided to the researchers at both the item and aggregate levels.
 - i. Each test answer is given in a separate column on one worksheet while answers are denoted as correct or incorrect on another.
 - ii. Aggregated data is provided for each student in columns giving number attempted, number correct, percent correct of attempted and percent correct overall for each test and subtest in the collection of pre- and post-tests.

- iii. Different versions of the tests with different answer keys and different numbers and types of question can be transformed by configuring tables in the database rather than by revising the code of the conversion program.
 - iv. Because this process is rapid, the researchers were able to divide the subjects into matched training and control groups on the basis of the pretest data for the fall 2004 experiment.
- b. We have nearly completed the program to provide a similar transformation of the data collected in the multiple-choice tests provided in the introduction module. In these tests, the student is immediately told whether her/his answer is correct and is required to try again if the first answer was wrong. Each attempt is a separate record. Timing data is collected for each attempt. For this data to be useful to the researchers, it must be converted to a form that gives the number of attempts and total time for each question, with aggregations across each set of questions.
 - c. We are developing programs to provide useful aggregations of data from the introduction and demonstration training modules. Much of this data involves timing of various actions. We are working out the format required for this data.

4. Extended reading comprehension tests

- a. We extended iSTART's on-line testing capability to include comprehension tests that had previously been delivered by paper and pencil. This is an important improvement because the on-line version eliminates the need for the transcribing of student answers. These tests are based on multi-paragraph readings that sometimes include graphics and employ both multiple choice and short answer questions.
- b. Computerization offered new possibilities of presentation and the opportunity to record the time the student spent answering questions and referring to the reading. After discussions and several prototypes we decided upon the following features:
 - i. The tests will include an instruction page, one or more reading pages, and a single page on which all questions will appear.
 - ii. The reading pages may have pictures.
 - iii. Test may be delivered with or without a time limit.
 - iv. The student may or may not be permitted to refer back to the reading pages once he/she has begun the question page.
 - v. A question appears only after the previous question has been visited.
 - vi. Answers may be revised.
 - vii. Student must click a button to begin answering a question in order to facilitate the timing of activities.
 - viii. Timing of all activities is reported to the database. This includes:
 - 1. Time spent answering each question and on each revision of a question.
 - 2. Time spent referring to reading pages.

3. Time reading the instructions. (The instructions are read before the clock starts running)
- ix. The database is updated frequently:
 1. when each question is answered
 2. when a page is visited, or
 3. every 10 to 15 seconds if there is no other activity.
- x. The test restarts where the user left off upon logging in after a session was interrupted by a crash or browser closing.
- c. Several tests have been built using the developed technology. Unfortunately, the current maturity of the system requires each test be hand coded according to a template or library and this process is prone to errors. We plan to generalize the technology so that no hand coding will be required. Instead, instructions, readings, and questions will be installed using an authoring interface.

5. Initiated reliable communication effort

- a. The realization in June 2004 that the fall 2004 experiments were to be conducted over wireless networks caused us to initiate a reliable communication effort. Our experiments over wired networks had resulted in very small and therefore tolerable loss of data, always less than 1%. Consequently we had taken no special steps to ensure that data was reliably collected in the server database. Wireless networks, especially those set up in an ad hoc manner as is common in classrooms, are less reliable than wired networks because they are subject to interference from other wireless networks and electronic devices such as cell phones. Consequently we began efforts to develop a more reliable communication between the iSTART server and its clients.
- b. The realization occurred too late to develop a reliable communication system for the fall 2004 experiments. However we are making significant progress on several fronts. These include a stop-gap improved reliability system, a reliable real-time communication system, a data recovery system, and a work-around for poor communication in the practice modules that require a response from the server to provide feedback to the student.
- c. Stop-gap improved reliability. In this approach, whenever new data is sent to the server database, the previous N data items are re-sent along with it, thereby increasing the likelihood that every data item reaches the server. This approach has been tested with the reading comprehension module. A variation on this method was already in use in the pre/post testing system which stored the entire sequence of answers each time the data was updated.
- d. Reliable real-time communication. We have developed a system for reliable real-time communication that is in the testing phase. This system is suitable for those modules that, after being downloaded, simply store information in the database on the server. This characterizes all the modules except the practice modules that require the server to provide the basis for interactive feed back to the user. So this can service all tests, as

well as the introduction and demonstration modules. This is reliable in the sense that all data will be sent to the server if the following conditions apply.

- i. The user is able to download the module to begin with.
 - ii. If communication is disrupted, communication is thereafter restored for a time sufficient to send the remaining data to the server.
- e. Data recovery system. This module stores data to the client's hard drive, allowing data to be recovered later. This system is under development. We tried several architectures that were unsuitable because of the security systems on Windows and Macintosh computers. We have developed an architecture that appears to be adequate to the task and are in the process of developing preliminary implementations.
- i. Disk-writing server. This module is a small http server that is contacted by any of the modules that are delivered to the client. The server writes the modules' data to disk from which it can be recovered later. The server approach had to be adopted because we needed to store data from multiple users of the same system in one place from which it can be retrieved under the security concerns of modern operating systems. When this module is completed, investigators will be able to recover the data from the hard drive but, without the development of the reporter module described next, will have to do so manually. This module can therefore be used before the rest of the system is developed as long as iSTART is delivered in an experimental setting.
 - ii. Reporter. This module resides on client computers and, when it is able, sends to the central database the data that has been written by the disk-writing server to the client's hard drive. The data sent might well have been collected from a user other than the current user of the system as would be typical in a computer lab environment. Work on this module will be completed after the module described next is finished since in an experimental setting the data can be recovered manually.
 - iii. Parser. Once a data file reaches the central database it must be parsed and a decision made about what to do with it. It is of particular concern that new data not be overwritten by old data, that duplicate data items not be entered as different data items, and that timing records be correct. Work has begun on this module.
 - iv. Work around for practice. The practice modules have special requirements that cannot be served with the systems being developed above. In particular, these modules rely on the server to provide the basis for feedback to the student being trained. The reliable communication modules being developed promise only to deliver the data at some time if there are communication difficulties. They cannot overcome network difficulties to produce a server response in a timely manner. Our solution is to build

limited intelligence into the client module and make our feedback less definitive when the limited intelligence is used. When the student supplies an explanation, it will be evaluated on the client at the same time as it is being sent to the server for evaluation. If no response is received from the server after a reasonable wait, the client evaluation will be used. Ultimately, the reliable communication system will be used to ensure that all responses, no matter which system evaluated them, are stored in the iSTART database. Responses not evaluated on the server will be tagged as such.

6. Investigation of multi-platform implementation.

- a. We hope eventually to be able to deliver iSTART training on the Macintosh platform as well as on Windows computers. All modules except the actual training have been constructed with this goal in mind and in many cases the testing modules have been delivered on the Macintosh platform during experiments. The major impediment to delivering the training on the Macintosh has been to find an alternative to Microsoft Agent technology. The alternative would be a cross-platform way of delivering characters that speak using a text-to-speech engine and that provide lip synching.
- b. While there are a number of systems for delivering speaking characters as described above, the only cross-platform systems with wide acceptance are Macromedia Flash and Shockwave. Unfortunately lip synching is not built into either platform. We discovered a shockwave add-in called DIRECTtts that provides text-to-speech capability and delivers the parameters appropriate to lip synching as speech is generated from text—e.g., the width and height of the mouth. We hoped that this could form the basis of a cross-platform system. However the vendor could not provide any example of anyone who had actually built a lip-synching system using their software and we discovered a prior customer who had tried to do so without success. We experimented with the DIRECTtts system and with several other approaches but concluded that attempting a conversion at this time would consume too many resources to be practical.
- c. Subsequent to reaching this conclusion, we found some indication that Macromedia may be providing lip synching capability in the future, so we have not given up hope.

7. Self-Explanation test module

One of our goals is to deliver all the assessment tests on the computer. This makes it possible to conduct all the phases of an iSTART experiment on computers as well as making all the materials online. We have started building a program that can deliver self-explanation tests on computer. These tests were previously delivered on paper and pencil generally along with the comprehension tests. After discussing and designing a couple of prototypes, we have developed a preliminary version of this test, version 1.0. This is being used in the experiments A, B and C being conducted in Fall 2004. The program allows us for flexibility, and has the following features:

- a. Gives the self-explanation test with texts that have multiple paragraphs.
- b. Allows us to predefine multiple tests for any group.
- c. MySQL database is used to store the tests. One of these is assigned an active status. This is the test the student will be given when they log into the program.
- d. The same test can be given to multiple groups if it's the active text assigned for that group.
- e. Allows us to select sentences for which the students need to self-explain.
- f. Version 1.0 displays sentences one by one. Sentences that require a self-explanation are displayed in Red. The current sentence is always bold.
- g. Allows us to display text in different colors, styles, font-styles, etcetera.
- h. Reading times are captured for each sentence. Typing time is captured when the student types a self-explanation. All the data is stored in MySQL database.
- i. The test is not given a maximum time for completion.

8. Weekly practice module

The texts selected by the teachers for extended practice are scanned and incorporated into iSTART as separate practice modules, after making a few minor adjustments. These adjustments include removing pictures, graphs, etc. because the current version of practice does not display pictures. The other step involved in creating a weekly practice is to assign a priority to the sentences in the text. The iSTART practice module uses this priority number to decide if it should follow-up the feedback with questions about other strategies. Currently this priority can be assigned easily using the Practice Authoring Tool. Our ultimate goal is to reach a stage where teachers add practice texts without any help from our team. They have made a first step in this direction. All the texts the teachers handpicked were made available as a library of texts on the iSTART server. Teachers go online to a webpage on our server and assign the text the class needs to practice that week.

9. Minor modifications to existing modules.

- a. Updating of test bank for pre- and post-tests.
 - i. New tests were entered into the test bank.
 - ii. Old questions were modified and updated.
- b. The pretest and posttest now identify questions left unanswered as well as questions that the student wishes to return to (the book-marking feature). A student who has questions remaining that were not answered is unable to proceed at the end of a test section until all questions are answered. Those questions are now much easier to find.
- c. The communication method between the client and the server was changed from using a 'slave window' to using an 'iframe'. While the strategy is conceptually identical, the slave window method created irritating occurrences when used on the Macintosh platform.

10. Expanding the pool of texts to be used for assessing self-explanations and comprehension

We have developed a pool of short texts relating to the domains of Life Science, Biology, Physical Science, Wellness, and Literature for the purpose of self-explanation

practice and comprehension assessment. Texts were selected by teachers (for the expanded self-explanation practice sessions) and the University of Memphis research team (for the pre-test, immediate post training test, and final post-test). We now have groups of (comparable) texts from a range of subject domains. Texts selected by University of Memphis team are comparable on several dimensions, such as cohesion and difficulty.

References

Best, R. M., O'Reilly, T. P., & McNamara, D. S. (submitted). Comprehending science texts: Effects of computerized reading strategy training. Paper submitted to the 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*.

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. (in press). 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.