Annual Report to the Institute of Education Sciences
August 15, 2005

iSTART: Interactive Strategy Training for Active Reading and Thinking (R305G040046)

Danielle S. McNamara (PI)

University of Memphis

I. EXECUTIVE SUMMARY

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

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

iSTART: Automated Strategy Training

iSTART is a web-based computer program that uses automated agents to provide reading strategy training. iSTART currently incorporates theoretically motivated Self-Explanation Reading Training (SERT; McNamara, 2004; McNamara & Scott, 1999), which teaches students to self-explain science texts by using active reading strategies known to facilitate and enhance comprehension, such as paraphrasing, elaborative inferences and bridging inferences (Gernsbacher & Hargreaves, 1988; Pressley, Wood, Woloshyn, Martin, King, & Menke, 1992; Rosenshine & Meisler, 1994).

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

Empirical studies with high-school and college students have shown that iSTART improves both reading comprehension and the quality of self-explanation during the process of reading (e.g., O'Reilly et al, 2004a, 2004b). The current IES grant aims to scale-up iSTART so that it can be more widely accessible to users of different abilities.

iSTART: The System We Envision

iSTART targets the kind of active reading strategies that research has shown to be characteristic of skilled, successful readers (e.g., Bereiter & Bird, 1985). These reading strategies are particularly important for understanding textbooks because they often include unfamiliar, challenging material. However, teachers are often inadequately prepared to incorporate reading strategy training into their classrooms (see, Snow, 2002). While they may be well versed in their particular domain, it is frequently the case that they have little training with regard to reading instruction, or how to incorporate reading strategy training into their classrooms (Baker, 1996; Snow, 2002). This impediment is augmented by the heterogeneity of students' needs (Cornoldi & Oakhill, 1996). Students each have particular reading deficiencies and learn reading strategies at different rates. Teachers clearly do not have time to give individualized reading strategy training to each of their students.

iSTART offers a solution to this educational problem by providing automated reading strategy training that is adaptable to students' needs and rate of progress through the program as well as to the teacher's course demands. However, to fulfill such an active role in classrooms, the current version of iSTART must be augmented in two principle ways: (1) by improving the ability of the system to adaptively respond to students' and teachers' needs and, (2) by creating a teacher interface that allows teachers to integrate iSTART into their classrooms.

Incorporating a computerized tutor into a classroom is not as simple as just giving it to the teacher and expecting that it be used consistently or successfully. There are many constraints that must be met in order for the system to be integrated into classroom in an effective manner. First, the teacher must understand the need for reading strategy training, and be receptive to intelligent tutoring systems. Second, the program must be easy for the teacher to use, and the system should also include a component that handles the questions about the program and data on the students' progress. Thus, a subcomponent called the teacher interface must be developed to facilitate teachers' use of iSTART in the classroom. Third, the number of course topics and range of text difficulty covered during the practice sessions need to be increased to make the system both usable in a variety of topics and appropriate to students of varying difficulty levels.

We are currently in year one of the three-year funding period. The summary below briefly describes our first year of progress toward our four objectives. Section II reports in greater detail the main objectives contained in the grant proposal and the progress made toward reaching these goals. Section III includes additional activities to refine iSTART, which are over and above the specific goals reported in the grant proposal. Overall, we have made good progress in terms of embarking on the goals pertained in the grant proposal and additional work that is critically related to the goals outlined in the grant.

Summary of Progress

Objective 1: Text Domains. The goal is to increase the number of texts used in the self-explanation practice. Also, by incorporating history and narrative texts to the existing science texts, students with varying ability levels will be able to use iSTART across three curriculum domains. We have gathered a large corpus of texts from different genres to add to the training corpora. We are also conducting experiments to examine the precision of the self-explanation evaluation algorithms across a wide variety of texts. One experiment with high-schools students was conducted last year to collect protocols from a variety of domains. We have also prepared an experiment with college and high-school students to examine self-explanations of science, history, and narrative texts before and after iSTART training. This experiment serves the dual purpose of providing additional protocols from a variety of domains (to refine our evaluation algorithms) and to test the transfer of reading strategy training with science texts to texts of different genres.

Objective 2: Adaptability to the User. The goal is to match the training to the level of the reader. Training will be tailored to the students with respect to their reading comprehension ability. Reading comprehension ability will be pre-tested using a battery of instruments, including general comprehension ability, metacognitive reading strategy knowledge, and science textbook knowledge. We will also vary the difficulty level of the texts students self-explain and the level of scaffolding provided in the practice section. We are working toward these goals through two approaches. First, we will attempt to improve the effectiveness and efficiency of the pre- assessment tools that capture the student's profile. Second, we will identify the difficulty level of the training texts (outlined in Objective 1) to match to the student's profile. Texts have been segregated into varying difficulty levels on the basis of grade level and text cohesion measures. The text cohesion measures are extracted from Coh-Metrix, a computer tool to measure linguistic and discourse features of texts (Graesser, McNamara, Louwerse, & Cai, 2004).

Objective 3: Responsiveness to Strategy Deficits. The goal is to expand the training modules so that we can match the reading strategy training to the level of the reader. For example, during initial training, a low-ability reader should master lower-level strategies, such as paraphrasing, whereas a high-ability reader should focus on learning higher-level strategies, such as constructing inferences through bridging or elaboration. We will therefore refine the reading strategy instruction so that readers' are taught strategies at the appropriate level. At this point, we have an extensive literature review on reading strategies that are theoretically and empirically supported and that could be incorporated into iSTART.

To make the system more user-adaptive, we will also refine the strategy training feedback (i.e., self-explanation feedback) so that it optimizes students' quality of self-explanations. For example, some students require extensive scaffolding to produce elaborations and bridging inferences (e.g., examples of strategy use), whereas other students may require little scaffolding (e.g., hints to elaborate). Once developed, we will evaluate the user-adaptable training by conducting experiments and usability studies.

Objective 4: Teacher Interface. The final goal is to create an interface to enable teachers to use iSTART effectively. The interface will (1) provide necessary information and training for teachers to use iSTART; (2) allow some degree of freedom for teachers to adjust the training program by selecting texts of their choice; and (3) gauge students' progress using the program. To date, we have developed prototypes for the teacher interface and conducted focus groups with teachers and students to evaluate iSTART, its implementation, and the teacher interface.

II. PROJECT GOALS AND PROGRESS

Objective 1: Text Domains

The goal is to increase the number of texts used in the self-explanation practice so that iSTART can be used across curriculum domains by students. Increasing text domains used in the practice session contributes to increasing the systems' ability to adapt to students with varying ability levels. The current version of iSTART provides self-explanation practice through two science texts, the development of thunderstorms and the origin of coal. While this has been sufficient to conduct our initial evaluations of the system, it is not sufficient to take iSTART to scale. First, increasing the number of texts is essential in order for struggling students to receive a greater amount of practice than currently possible. Although practice with two texts provides substantial benefits, our research has shown that this is not enough practice for the students most in need of training. Second, our experiences with students and teachers over the past few years have made it evident that the reading strategy training offered by iSTART should be available in other domains. The topics of history and literature have been consistently listed over half and up to 75% of the high-school SERT participants in our studies as topics for which they have used SERT strategies (though they were trained with science texts).

Expanding the number of topics covered by iSTART is beneficial for both applied and theoretical reasons. First, teachers will be more receptive to using iSTART in their classroom if topic-relevant material (or better yet, selections from the course textbook) is available for the student to read on the iSTART system. iSTART will be better used by teachers if it allows coverage of required content while their students are learning the reading strategies. In theoretical terms, increasing the number of topics covered will contribute to our understanding of reading processes. It will both afford us the opportunity to examine how different strategies are used and also contribute to comprehension across different domains.

Our research plans to meet Objective 1 are represented as sub-goals in Table 1. The table indicates progress made with each sub-goal and the time frame in which we expect to complete the work.

Table 1: Sub-goals for Objective 1

Sub-goal	Description of activity	Progress	Time frame
1.1	Create a library of digitalized texts	Commenced	Years 1 and 2
1.2	Create a self-explanation evaluation system that can be easily adapted to a wide range of texts	Commenced	Years 1 and 2
1.3	Evaluations of the feedback system using Self-explanation protocols of narrative text collected in 2004 (Study 1)	Commenced	Years 1 and 2
1.4	Experimental study involving collection of self-explanations for science, history and narrative texts	Planning phase	Years 1 and 2

Sub-goal 1.1: This goal involves building text libraries for iSTART practice by increasing the number of texts available in the self-explanation practice so that iSTART can be used more extensively both within curriculum domains (e.g., different topic within science) and across curriculum domains (e.g., science, history, and literature). At this point we have acquired a large corpus of texts from different genres. We segmented several high-school texts from the domains of history, science and literature into meaningful units suitable for training purposes (about 400 words or 20 to 30 sentences). This produced approximately 400 samples of text for each of the history, science and literature domains. Currently we are evaluating the samples in terms of topic difficulty, document structure, and sentence complexity.

Sub-goal 1.2: Goal 1.2 involves the refinement of a self-explanation evaluation system. This system gives rapid evaluations and responses to student self-explanations. It also, can easily be adapted to a wide range of texts. The self-explanation evaluation algorithms must be easily applied to a wide variety of texts. Our first version of an algorithm to achieve this goal is described in McNamara, Boonthum, Levinstein, and Millis (in press). We have created an evaluation system that can be automatically applied to virtually any text and is highly predictive of self-explanation quality. This revised system was used in the experiments conducted during the 2004-2005 school year. Informal examinations of the accuracy of the system's evaluations of self-explanations over a variety of texts indicate that the system performs quite well. We are currently completing the human coding of the protocols, which is necessary for a systematic evaluation of the system (described in sub-goal 1.3). We will revise the system based on the analyses of these protocols.

We are also working to create new semantic spaces for the LSA algorithms that are part of the system. For example, as we need to add a history space, we must gather a large amount of history texts. The spaces contribute to evaluations by providing a semantic comparison of the self-explanation to the target text. We have thus far created an LSA space for narrative texts, and are in the process of developing the history space. Unfortunately, preliminary evaluations of the narrative space have indicated that it does not perform well (i.e., it does not produce results equivalent to or as good as other spaces). As such, we are improving this space by expanding it and combining it with other, smaller narrative spaces. We have also gathered texts to create a history space, which will be compiled and formatted this fall. We will compare the results using separate and combined spaces, though our hope is to be able to use one large, combined space for all texts.

Sub-goal 1.3: This sub-goal involves the collection of and human coding of protocols across a wide variety of texts to support the revision and evaluation of the system's algorithms. This year, we completed a preliminary study in which high-school students self-explained four types of texts (science, literary analysis, narrative, and wellness) after receiving iSTART training with science texts. The protocols from this study are currently being rated in terms of strategy use and quality by human raters. These ratings will be compared to the judgments made by the system (McNamara et al., in press; McNamara et al., 2004; Millis et al., 2004). The match between trained raters' judgments and the system's judgments (in terms of d-primes and Cohen's Kappa) indicates the appropriateness of the judgments made by the system. Half of these protocol ratings will be used to train (revise) the system and the other half will be used to test the system's effectiveness. In addition to the protocols collected last year, we will also use protocols collected in the experiment described below, to be conducted this fall.

Goal 1.4: We are preparing an experiment in which we test the transfer of self-explanation training on science texts to texts from three domains (science, history, and narrative texts). This experiment also allows us to collect additional self-explanation protocols from history and narrative domains. The experiment uses a within-subjects design such that participants will produce self-explanations for science, history, and narrative texts before and after iSTART training. For each text domain, we will collect self-explanation protocols from 54 high-school students and 54 college students (i.e., N=108). Prior to training and testing, individual difference measures will be collected to assess prior world knowledge, prior domain knowledge, reading comprehension abilities, and reading strategy knowledge.

At this stage, we do not have training materials specific to history and literature. Therefore, participants will be trained on science texts, but tested in one of the three domains. This study serves two goals. First, the study will provide data regarding the extent to which the effect of self-explanation training in one domain or text genre transfers to another domain or text genre. Second, the protocols collected will allow us to fine-tune the algorithms used to evaluate strategy quality. This is an iterative cycle in which we develop computational linguistic algorithms, compare the output of these algorithms to human ratings for a subset of the protocols, refine the algorithms, and then reexamine the algorithms using the remaining protocols. The work to prepare this experiment involved creating measures of prior knowledge appropriate for both college and high-school students, and selecting appropriate texts (2 texts for each domain) by considering various relevant dimensions (e.g., word frequency, grade level, cohesion, etc.). We will begin the experiment in September.

Objective 2: Adaptability to the User

It is essential that we increase the range of difficulty of the training texts in order to serve a wider range of students. The current training texts are targeted at high-school students. However, for some readers, these texts may not be challenging enough to stimulate processing at the zone of proximal development (Vygotsky, 1978). Our past studies with high-school students indicate that students vary greatly in terms of their reading skills, including their ability to self-explain texts using reading strategies. Whereas some advanced students can use elaboration based on general world knowledge and personal experiences, many students often struggle with producing paraphrases when a target sentence is long, complex, and or contains several unfamiliar words. Objective 2 is aimed at providing practice that adapts to students' current ability level. Our research plans and progress for meeting Objective 2 are represented as sub-goals in Table 2.

Table 2: Sub-goals for Objective 2

Sub-goal	Description of activity	Progress	Time frame
2.1	Increase range of text difficulty	Commenced	Years 1 and 2
2.2	Matching the text to the reader (Study 3)	Not commenced	Years 2 and 3

Sub-goal 2.1: The texts contained in the text library described earlier will be analyzed and indexed in terms of various relevant text features that affect difficulty level. We have analyzed high school science, history and narrative texts (see goal 1.1, Objective 1) using Coh-Metrix (Graesser McNamara, Louwerse & Cai, 2004; McNamara, Louwerse, & Graesser, 2002; http://cohmetrix.memphis.edu). Coh-Metrix provides measures of text characteristics indicative of various types of difficulty (e.g., word frequency, sentence length, and syntactic complexity, and cohesion), that helps us compile and organize an effective text library. We are currently in the process of determining which of the many Coh-Metrix measures will be used to index the text difficulty in the library.

Sub-goal 2.2 (Study 3): In Years 2 and 3, we will address the question of matching text to reader. We plan to achieve this through a study in which the difficulty of the practice text (in terms of cohesion) is manipulated at three levels (low, moderate, high). This study will include 240 college and high-school students varying in terms of reading ability and domain knowledge. Following training at one of these three text difficulty levels, participants' self-explanation ability will be examined using a text of moderate difficulty. This experiment will afford examination of performance and improvement during training as a function of reader aptitudes, as well as the generalization of training to the moderate difficulty target text. We expect to observe an interactive effect such that low-ability readers will show greater improvement during training using the high-cohesion text, and high-ability readers will show greater improvement using the low-cohesion text (see e.g., McNamara, Kintsch, Songer, & Kintsch, 1996). We expect similar trends in terms of transfer of training to the moderate difficulty text.

Objective 3: Responsiveness to Strategy Deficits

Objective 3 is designed to make iSTART more adaptive to the level of the student. This goal, in part, will be realized by increasing the range of texts used in the practice. However, achieving this objective requires substantial modification and improvement in the training system. Specifically, four aspects of the system require further work: 1) refine the pretesting procedure that serves as a foundation to build a student model to guide the training process (e.g., assignment of text, strategies, etc.); 2) increase the range of strategy instruction provided; 3) further modify and improve the feedback system of the practice module such that the system will be capable of providing appropriate feedback to students with varying ability, and dealing with texts from a variety of content domains (narrative, history, science); and 4) creating a teacher interface that enables high school teachers to interact with iSTART effectively when using the system in the classroom. The progress made in each of these three areas is reported below. The tasks associated with Objective 3 are outlined as sub-goals in Table 3.

Table 3: Sub-goals for Objective 3

Sub-goal	Description of activity	Progress	Time frame
3.1	Refine battery of individual difference measures	Commenced	Years 1 and 2
3.2	Create and use a student model of strategy needs	Commenced	Years 1 - 3
3.3	Expansion of strategies included in the system	Commenced	Years 1 and 2
3.4	Usability assessments (Study 4)	Piloting commenced	Years 1 - 3
3.5	Assessing the efficacy of the expanded iSTART system (Study 5)	Not commenced	Year 3

Sub-goal 3.1: In previous research we have assessed students' competencies and aptitudes using a large battery of pretests prior to training using iSTART. The pretest measures include general reading skill (e.g., Gates-MacGinite reading test), metacognitive reading strategy knowledge (Schmitt, 1990), metacognitive skills (e.g., Mokhtari & Reichard, 2002), general science knowledge and specific science vocabulary knowledge. Over the past few years, we have collected data on thousands of participants that use our existing battery of pretest assessments. This large data set enables us to evaluate the ability of our measures to predict student performance on students' ability to learn from and comprehend texts before and after training using iSTART. Our efforts have focused on evaluating the validity of our pretest measures in terms of predicting students' performance on 1) science passage comprehension; and 2) self explanation quality produced for science texts before and after training using the iSTART system.

While it is acceptable to conduct two-hour pretest sessions in experimental situations, it would be impractical to expect teachers to follow a two-hour pretest session to determine the appropriate level of training. Our goal now is to maximize the pretest performance while eliminating redundancy and reduce the time necessary to administer these tests. Since the start of the funding period, we have evaluated each pretest measure and constituent questions in terms of whether they predict students' comprehension performance on a variety of science texts and self-explanation scores. Items (or entire measures) that show little or no predictive validity are marked and considered for elimination from the battery whereas measures and questions that indicate a robust ability to predict comprehension over a wide variety of science passages are retained.

Using the methods described above, we are currently in the process of reducing the amount of pretest assessments, while preserving or even improving the accuracy of the assessment relevant for designing and tailoring the training program. While we have not completed our final evaluation of the assessments, we have been able to successfully identify questions with a high level of predictive validity. For example, we have successfully reduced our general knowledge measure from 35 questions to 20 questions while retaining almost the same overall predictive validity. We have also been able to reduce 20 minutes of testing time by identifying and possibly eliminating two of our current tests due to either redundancy or their overall poor predictive validity. In short, we are confident that we will be able to reduce our current assessments while maintaining a similar level of predictive power.

Sub-goal 3.2: Our second goal is to create and use a student model of strategy needs. The model will be based on data gathered from the pretest battery and from the student interactions in the course of training. We conceive a dynamic student model that makes use of pre-assessment data to initiate training and then continuously updates as a function of training performance data, such as response times and question answering accuracy (e.g., strategy quiz performance). For example, in the introduction, the student is given a brief quiz after each strategy is introduced. In the demonstration section, we learn which strategies the student is able to, or prefers to identify. In the practice section, we learn how well the student is able to explain the texts and to identify strategies that are used. As this information is added to the student model, it will be possible to predict which avenues of iSTART training are most valuable as the student progresses.

We intend to make use of theoretical guidelines as well as empirical data to direct training. Both theory and data will guide options such as what level of scaffolding to provide, what strategies to emphasize, the number of examples to provide, and the amount of practice to offer. Constructing and revising a model of the student throughout training will allow us to map the characteristics of the student to an appropriate program of training. To this end, we will use data collected from previous experiments to examine the link between reader aptitudes (pretest assessments mentioned above) and the success with which students' learn and use strategies incorporated in iSTART (e.g., paraphrasing and bridging). Specifically, we aim to identify how students' level of reading skill and prior topic knowledge is associated with students' performance on the iSTART module tasks (quizzes incorporated in the Introduction module) and use of strategies (self-explanation quality rating generated in the practice module). The overarching idea is to determine which reading strategies are appropriate for students of differing ability levels. For example, students with poor reading skills may demonstrate a poor knowledge of higher-level strategies (e.g., elaboration and bridging) and thus benefit more from learning the lower-level strategies (e.g., paraphrasing). Conversely, students with high reading skills may demonstrate a better grasp of the higher-level strategies and thus benefit from higher-level reading strategy training.

Sub-goal 3.3: One approach to improving the system's responsiveness to the user will be to increase the range of strategy instruction, including modules providing instructions on more basic reading strategies on the front end, and high-level reading strategies on the back end. These additional modules will augment the system's ability to adapt to the needs of the student. For example, our research has indicated that some students need more training on paraphrasing before continuing on to the higher-level strategies. Although paraphrasing does not lead to deep comprehension, it is the backbone of self-explanation. Some readers need a greater amount of scaffolding to learn successful paraphrasing. This scaffolding will include more information on what it means to read, understand, and learn from text, how to discern the meanings of words, and how to transform a sentence into familiar terms.

We recognize that the range of metacognitive reading strategies is broader than the strategies taught by iSTART. Including new strategies to the repertoire of iSTART needs to address several issues: 1) add training appropriate to students who have progressed successfully through the iSTART curriculum, 2) provide training to bridge deficiencies that students may suffer that prevent them from being successful with iSTART, and 3) to provide alternative curricula that may better suit individual students' learning styles.

Expanding strategies in iSTART requires an extensive literature search on appropriate strategies, understanding and evaluating the strategies, and finally analyzing the relations between different strategies. To this end, we have conducted an extensive literature review of existing reading strategies that are not currently a part of iSTART. Our focus is investigating empirically supported interventions in the literature. We are in the process of selecting the most effective and feasible interventions by approaching this task from the two directions: 1) development of reading strategy standards; and 2) hosting a conference on reading strategies.

While the work with the College Board is not funded by IES, it is relevant to achieving our goals in this project. In collaboration with the College Board, we are in the process of developing reading standards for comprehension from grades 6 to 12. This endeavor has two goals: 1) to identify the specific elements of reading skills required to master reading comprehension at different stages of development (i.e., grade level); 2) identify interventions designed to teach and improve the skills necessary to achieve the level of reading proficiency that are appropriate for the particular grade level in question. Our work centers on the development of reading strategy standards, which identifies the most important strategies to be mastered by middle-school and high-school students. Our work on these standards will also serve to guide our choice of additional strategy interventions to incorporate into iSTART.

We also hosted a conference on reading strategies that encompassed researchers on reading comprehension research from a variety of disciplines (psychology, education, and computer science) and countries. The talks at this conference will be submitted as chapters for a book on reading strategies to be published by Erlbaum (McNamara, in progress). The focus of the conference was to discuss and exchange theories and findings relating to reading comprehension interventions. Overall the conference was a great success: a great deal of cutting edge research on reading comprehension training was discussed among internationally renowned experts on readings strategies. The result of this conference was an invaluable experience that will continue to shape our plans for iSTART.

Sub-goal 3.4: Usability testing will be conducted to test the revised iSTART system. Our approach to usability analyses is three pronged, using a combination of information gathered via think-aloud protocols, eye tracking, and task-completion rates. While the usability studies are primarily planned for years 2 and 3, we have conducted one pilot investigation to tap into students' eye-tracking movements of college-student users when using the program. The purpose of the eye-tracking experiment was to investigate where students focused their attention while working with iSTART. For this experiment, investigators used a modified version of the iSTART program which included the introduction and demonstration sections.

Eye-tracking videos were collected for eight college students using an SMI iView Hi-Speed Eye Tracker. Participants interacted with iSTART for approximately 45 minutes while their eye movements were recorded onto a DVD using the scene video eye-tracking option, DVD recorder, and television. Eye movements were indicated using a circular gaze cursor which increased and decreased based on participants' eye gaze duration.

Of particular interest is the utility of the animated agents in iSTART. Animated agents deliver the training to make use of interactive dialog with the trainee, and to render training more interesting. One question is whether these agents are engaging or distracting. When we measure the times that the participant looks at the agent in relevant situations, such as when the agent speaks, we can compute hit rates and false alarms. Different predictions can be made accordingly: 1) if agents in general are of no interest to the user, we expect to find low hit and false alarm rates (and low d primes) for all agents because the participant will randomly look at the agents; 2) if a particular agent attracts a student's interest, but for other reasons than entertainment (e.g. the agent looks peculiar or funny), hit rates for that agent will be higher than for other agents, but false alarms will also be high (resulting in low d primes); 3) if the agent attracts the student's attention at the right time, hit rates will be high, false alarm rates will be low and d primes will be high accordingly for all agents. Four areas of interest were identified on the screen: the main character (Dr. Julie) and two supporting characters (Mike and Sheila), the iSTART logo and the "next screen" button. The three animated characters were considered the main areas of interest for the purposes of this study.

Hit rates were computed as the proportion of times the participants correctly fixated on the agent, false alarms as the proportion of times the participants fixated on the agent during non-events. Overall d prime scores on the agent were high (2.66), with high hit rates (.85) and low false alarm rates (.05). This suggests that the participants look at the agent when the agent speaks, but does not look at the agent when the agent does not speak. These results were found for all three agents. Thus, when the agent talks, participants pay attention to the agent, but they do not look at the agent when the agent is quiet. The exception to this is that participants move back and forth between text balloon and agent while the agent is talking. We also found that participants pay attention to the relevant agent when the speaker refers to that agent.

Objective 4: Teacher Interface

As yet, iSTART has been implemented in classrooms solely under experimenter control. Scaling up the program means that the classroom teacher would control when and how the program is used. Thus, Objective 4 is to design and assess a teacher interface so that teachers can implement iSTART in classroom settings. Our research plans and progress to meet Objective 4 are represented as sub-goals in Table 4.

Table 4: Sub-goals for Objective 4

Sub-goal	Description of activity	Progress	Time frame
4.1	Development of teacher interface	Commenced	Years 1 – 3
4.2	Teachers' needs assessment (Study 6)	Pilot study conducted	Years 1 – 3
4.3	Assessment of teacher interface (Study 7)	Not commenced	Year 3
4.4	Assessment of teacher interface in classroom (Study 8)	Not commenced	Year 3

Goal 4.1: We envisage a teacher interface that comprises three modules (outlined below) Progress made toward the development of the teacher interface is discussed after the modules.

(1) Teacher Instruction Module

As mentioned above, we contend that teachers need to have an understanding of psychological theory underlying the iSTART system. We provide basic information about the reading strategies and why students need them. This information will cover topics such as theories of text comprehension, the effect of prior knowledge on reading, and the importance of reading strategies. This will promote teachers' understanding of the reading process and understanding texts. The teachers should also have the opportunity to have "a hands-on" tour of the modules to get an idea of the tasks their students will experience. We can also provide a tour of the features in other modules and how to use them.

(2) Training Organizer and Manager

The goal of this module is to provide technical and practical support to the teachers so they can use iSTART effectively. The option of being able to customize the training towards various ability levels of students is critical in making the iSTART training more effective. The teachers will be able to designate different training sequences or modules for different groups of students in the same class or different classes. The program will have default sequences for students based on their assessed ability; however, this module will allow the teacher to modify those sequences. The teacher will also be able to choose training and practice texts. We will provide a library of texts that will encompass a wide selection of topics and domains from which to choose. The library will also contain texts teachers need/use in their classroom.

(3) Performance Analyzing Tool

The purpose of this module is to provide a means to monitor the students' performance and progress in the training. The tool will provide reports concerning the assessments given prior to iSTART training and reports on students' performance during training. These reports will help teachers gauge the depth of difficulties students are experiencing. This tool will also help teachers create an iSTART training sequence that will be more suited to help the students in areas most in need of improvement by providing information about student participation and progress. Although these reports present summary information, the teacher will have access to more specific details such as modules completed, time spent, number of attempts, and performance on these attempts.

The design of the teacher interface requires a good deal of preliminary exploration. Each module presents unique challenges which need to be dealt with accordingly. The Teacher Instruction Module involves challenges such as determining the appropriateness of the instructional content, the extent of content that teachers need to understand, the format in which the material should be presented, and whether the material should be designed for the teachers to go through by themselves or with the help of an expert teacher (in a workshop). The issue of usability is of critical importance for the Training Organizer and Manager module. This is due to the fact that the teachers will be using this module primarily to moderate the iSTART training for their students. Therefore, the main challenge in this module is to conceive a design that is easy to use and allows teachers to complete all the steps needed to set up and customize the training for their students. In the Performance Analyzing Tool Module, the main challenge involves determining the right amount of data to be included in the reports generated for the teachers.

Prototype designs have been created to examine the feasibility of several approaches. Prototype development has involved consulting with teachers and exploring published research. We interviewed teachers who participated in the iSTART experiments conducted this year. This allowed us to gain a better understanding of issues involved in designing software to be used by the high-school teachers. Some conclusions based on these interviews are discussed below.

Goal 4.2 (Study 6): One of the critical steps toward integrating iSTART into the classroom involves obtaining feedback from the teachers. iSTART is to be developed in a way that enables teachers and students to use the tool with ease and efficiency. As a first step toward assessing teachers' preferences when using the interface (e.g., the level of information supplied), we conducted a pilot study in which previous teacher-users gave their opinions about iSTART and the teacher interface. The opportunity to interview teachers developed out of four experiments conducted within the original NSF grant (entitled Promoting Active Reading Strategies) in which four teachers led high-school students through iSTART in real world classroom settings.

We asked teachers about the usability and effectiveness in promoting students' comprehension ability and general class performance. For example, we asked questions such as "Explain any difficulties you encountered administering the training?", "Do you think the students benefited from the training in terms of how to better understand a text?", and "Do you think the students enjoyed the training?" The interviews also provided an opportunity to explore the question of how we should design the teacher interface. For example, we asked teachers whether they would like to have a freedom of using their own texts in the Practice module.

We obtained some valuable suggestions about the instructional content and how to make it more effective. The teachers also provided some insight on the challenges faced by teachers when they try to use intervention programs such as iSTART.

The development of prototypes and design templates has allowed us to present possible designs, simulate discussions and obtain feedback from both psychologists and teachers. This process has led to the conclusion that detailed information presenting the scientific basis of iSTART, measures of success, and so on, should be available to the teachers. However, the teachers also need the option of accessing either an abridged or detailed version. Based on the suggestions provided by the teachers we reached a similar conclusion about the Performance Analysis Tool; that is, teachers need to be able to opt between a summary report and a detailed report of their students' performance. The teachers indicated that once they identify the students who need help based on the summary reports, they will need the ability to look at the more detailed charts. This will enable them to more accurately understand the problems faced by that particular student and try to address them. Also, teachers' recommendations confirmed most of our initial design assumptions about the Training Organizer and Manager module. We assumed that teachers would like to have the ability to tailor the number of reading strategies and quizzes in the training phase. Teachers' feedback indicates that most teachers prefer a hands-on tour of the training program to obtain an idea of the course of the students' training. Along these lines, teachers also have indicated that other features, such as ability to upload their own texts and the ability to reuse other teachers' texts, will encourage teachers to use iSTART more often.

Goal 4.3: We plan to conduct usability studies with the teacher interface during Year 3. We will use a structured series of formative evaluations including one-to-one and small group sessions with teachers. Please refer to the grant proposal for details.

Goal 4.4: In Year 3, we also plan to have a user interface and iSTART system that can be integrated within a limited set of classrooms (i.e., approximately 30 classrooms). This system will be tested at a small scale with 6 teachers within a single high-school in Shelby County, Tennessee. Please see the grant proposal for details.

III. Additional activities to refine iSTART

Activity 1: Student views of iSTART

To successfully develop and integrate iSTART into the high-school classroom, it is critical that we obtain both positive and negative views of students who have trained on the system. Student views provide important insight into 1) aspects of the system that are helping students learn to use the iSTART strategies, and 2) aspects of the system that may hinder reading strategy learning (e.g., problems with the characters and the understandability of the instruction offered in the Introduction module). Importantly, aspects of the system that students dislike or find difficult need to be treated seriously. Students will not fully benefit from training if the strategy instruction is too complex or the feedback given in the practice module is frustrating. We must use student opinions to reconcile differences in what we (the researchers) believe to be optimal methods of reading strategy training and students' views of optimal reading strategy training.

As a first step to identifying the ways in which the current system (iSTART v2.0) needs to be refined to meet the needs of students, we have recently conducted focus groups with high-school students about their views of iSTART. The opportunity to conduct focus group discussions developed out of experiments conducted with ninth grade students that formed part of the original NSF grant (entitled Promoting Active Reading Strategies). Following the experiments, 14 focus group discussions were conducted with a selection of students trained with iSTART. Each focus group comprised five ninth-grade students of mixed ability (as defined by performance on the reading strategy knowledge questions incorporated into the iSTART Introduction module). Using a semi-structured interview, we asked questions pertaining to general themes (positive aspects of iSTART; negative aspects of iSTART; views about the animated characters, views about Merlin's feedback and suggestions for improving the system).

We are currently processing information elicited from students in the focus group interviews. Our focus is to identify issues recurrently surfacing among different groups and determine how frequently the issues were raised (i.e., how many of the groups raised the issues). Table 5 provides a summary of the preliminary analysis from all focus groups; that is, the table organizes various recurrent issues which surfaced among members of the focus group in terms of five themes, and show how frequently (number of focus groups) these issues surfaced out of a possible 14.

The data from theme 1 indicate that most groups of students believed iSTART is helpful in terms of learning reading strategies. Specific comments indicate that many students found iSTART cognitively challenging, and that they invested a considerable amount of effort when participating in the practice components. Theme 1 data also suggests that many students think the system was "pitched" at an appropriate difficulty level. In terms of the negative aspects (theme 2), many groups reported that students became bored during the extended practice sessions and some commented that the texts were not interesting. Moreover, comments made by some students indicate that some practice texts were difficult to understand because the vocabulary was too difficult. It should be noted, however, that the practice texts during the extended (extra) practice sessions were chosen by the teachers and excerpted from the students' textbooks.

Table 5. Themes emerging from the focus groups

Theme	N
1. Positive aspects of iSTART	
Helped learn reading strategies	7
System was well designed	7
Strategies taught at an appropriate level of difficulty	8
Enjoyed using laptop computers	4
2. Negative views of iSTART	
Became bored during extended practice	9
Found the texts boring	6
3. View of animated characters	
Voices were poor quality	7
Characters were childish	6
4. Views of Merlin's feedback	
Frustrated in practice (prompts to retype "adequate" self-explanations)	13
Found Merlin's feedback useful	2
5. Suggestions for improvements	
Make the practice less demanding (don't push students so much)	2
Shorten the training/practice sessions	12
Use different practice texts (more interesting)	3
Incorporate option to ask questions/add dictionary	2

As for the system design (themes 3 and 4), approximately half the groups mentioned that the character voices were poor in quality, and the animated agents (e.g., Merlin and Genie) were childish or immature. Importantly, all but one group voiced their frustration about the content of Merlin's feedback in the practice module. Specifically, they thought Merlin asked for more information to be added when the students believed their self-explanations were sufficient. (theme 4). More specific comments indicate that most participants were prompted to produce multiple self-explanations for many of the sentences contained in the practice texts.

Finally, in terms of improvements to the system (theme 5), the most prevalent suggestion was to reduce the amount of time spent on individual training sessions. Currently, students spent entire class periods going through the system (about 45-50 minutes). An analysis of the specific comments indicate that students are willing to participate in regular training and extended practice, although they would prefer to spend less time during each individual session. The students suggested 20 to 30 minutes would be preferable. Shorter, more focused sessions would increase engagement with the activity.

Overall, the preliminary analyses suggest that students feel positive about the usefulness of iSTART as a reading strategy training tool. There are, however, clear concerns with regard to their engagement with the system (theme 2) and frustration about Merlin's feedback (theme 4). Two implications emerging from the focus groups are that 1) Merlin's feedback should be less demanding in some cases, and 2) individual iSTART sessions should be reduced to a maximum of 30 minutes.

Activity 2: General improvements to the iSTART system

Over the past year, we have made several improvements to iSTART that are of critical relevance to the goals of the IES grant. These improvements are classified as: 1) refinement of the demonstration module; 2) refinement to practice module; 3) creation of a modified practice text authoring tool; and 4) improving the communication between the server and client computer. Below we describe the work that has been done in these core areas.

1) Refinement to the demonstration module

We conducted an analysis of the students' performance on the iSTART demonstration module beginning with a comparison of the performance of experiments conducted in 2004 with the performance of the Old Dominion college students in spring 2002. We found that although the design of the module accorded well with the performance of the college students, it was ill adapted to the high-school students. The module was designed to provide a gentler questioning approach to students who were performing poorly and a less instructive, more testing, approach to those performing well. The college students moved in both directions with equal frequency, but the high school students overwhelmingly moved toward worse performance. We concluded that the task in demonstration is too difficult for most high-school students, and invented a variety of methods to simplify the task for them. The methods include 1) focusing the student's attention on just a portion of the self-explanation under examination, 2) visually breaking the self-explanation into parts instead of asking the students to do their own parsing, 3) adding reminders (tool tips) to the multiple-choice box that is used to question the students, and 4) incorporating more pedagogy into the interaction by telling the students the right answer when they get something wrong and by asking them to justify a correct answer. We are planning to incorporate these changes into the module but need to perform some preliminary analyses that will allow us to construct a scale of difficulty for various combinations of these technologies.

2) Refinement to the practice module

During previous experimental studies assessing the efficacy of iSTART as a classroom teaching tool, we observed that the students became frustrated when they had to self-explain each sentence in a text, especially when they were engaged in the weekly practice sessions (See Activity 1, Section IV). This suggested that an alternative approach, where only some sentences were required to be explained, would be better received. In support of this approach it should be noted that self-explanation is *not* meant to be used on every sentence—the "explain every sentence" approach is actually a somewhat unnatural use of self-explanation, one adopted to provide concentrated training in a short time. Consequently a new Practice module was designed and implemented. This one allows an administrator to specify a subset of the sentences as target sentences to be explained. This module currently uses the same algorithms to provide feedback as do the regular practice modules. Since we have modified the practice module so that students explain selected sentences rather than every sentence, we have to reexamine the formulae we use to evaluate their explanations. Explanations that occur every few sentences are generated in a broader context and the student should incorporate the material appropriate to the intervening sentences into the one actually produced. Consequently we need to re-examine what constitutes good, fair, and poor explanations and revise our formulae to better recognize them. In the following year, we will examine the explanations produced to determine whether the algorithms or benchmarks should be revised due to the modification.

3) Modified Practice Text Authoring Tool

To accommodate the new target-sentence version of Practice just described, we modified the Text Authoring Tool such that the experimenter or teacher can specify which sentences are target sentences to be explained. The JavaScript file that is created by the tool for the text can be used in all versions of the practice module (All-Sentences with and without feedback and Target-Sentence with feedback). This change also required some changes to the underlying data structure.

4) Improving the communication between the server and client

To run iSTART in classroom settings, it is advantageous to have wireless networks to achieve a reliable communication effort. It would be impractical to hardwire computers in real world classroom settings. Further, it is safer to use wireless systems in the classrooms as hard-wiring exposes students to many dangers, such as loose wires. Our previous experiments over wired networks had resulted in very small and therefore tolerable losses 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. We were subsequently made aware that the experiments were also dependent on school networks which were often quite unreliable. School computers were found to be often infested with malware (i.e., ***mal****icious* *soft****ware***, designed specifically to damage or disrupt a system, such as a virus). Consequently, we began efforts to develop a more reliable communication between the iSTART server and its clients. The realization occurred too late to develop a reliable communication system for iSTART experiments conducted in 2004. However, we have made 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.

- **Stop-gap improved reliability**. In this approach, whenever new data is sent to the server database, the previous N data items are 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.
- **Reliable real-time communication**. We have developed a system for reliable real-time communication that has been tested and is ready for incorporation in the appropriate iSTART modules. This system is suitable for those modules that, after being downloaded, simply store information in the database on the server and do not require continuous or frequent contact with the server in order to interact with the user. This characterizes all of the modules except the practice modules that require the server to provide the basis for interactive feed back to the user. So this can support all tests, as well as the introduction and demonstration modules. This module maintains a queue of data to be sent to the server and verifies that each item reached the server before removing it from the queue. 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.
- **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. These systems made it difficult for web-based applications to access the user's hard drive, segregate data, and access data by user. We have developed an architecture that appears to be adequate for this task and have completed part of the system.
    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 the data 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 such that the data can be retrieved according to the security criteria of modern operating systems. This module has been completed and once it has been tested, investigators will be able to recover the data from the hard drive. However, without the development of the reporter module which will be described next, the retrieval needs to be done 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. This module is currently under development.

iii. **Parser**. Once a data file reaches the central database it must be parsed and a decision must be 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.

iv. **Policy Issues**. Since this system must be installed on the user's computer, the system administrator should be notified of, and agree to, the installation. We need to decide whether to require this module's installation as a prerequisite to using iSTART. We also need to determine how to make this module compatible with firewalls and anti-spyware software.

- **Work-around for interaction**. In the practice module, the intelligence that evaluates the students' protocols in order to provide coaching feedback is located on the server. Therefore this module requires continual contact with the server: the student's protocol is uploaded and the server's evaluation is downloaded. The reliable communication system described above will not work in this environment in any form since that system uploads information after a delay of what might be seconds, hours, or days and the coaching must follow immediately on the student's response. In dealing with network problems this module can either notify the student that it is unable to function or it can substitute local intelligence for the intelligence that resides on the server. The local intelligence would be provided by using a modification of the system used in the first version of iSTART, which was a JavaScript application whose ability to predict the quality of a student protocol was considerably lower than the intelligence in the current version. We have not yet decided whether to incorporate this work-around into iSTART, but we have the basic design for including it: the protocol would be sent to both the server and the local system. If the server value does not return within a set time, the local system's response would be used.

References

Baker, L. (1996). Social influences on metacognitive development in reading. In C. Cornoldi & J. Oakhill (Eds.), *Reading comprehension difficulties: Processes and intervention* (pp. 331-352). Mahwah, NJ: Earlbaum.

Bereiter, C., & Bird, M. (1985). Use of thinking aloud in identification and teaching of reading comprehension strategies. Cognition and Instruction, 2(2), 131-156.

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.

Cornoldi, C., & Oakhill, J. (Eds.). (1996). *Reading comprehension difficulties: Processes and intervention.* Mahwah, NJ: Erlbaum.

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

Graesser, A., McNamara, D., Louwerse, M., & Cai, Z. (2004). Coh-Metrix: Analysis of text on cohesion and language. *Behavioral Research Methods, Instruments, and Computers, 36*, 193-202.

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

McNamara, D.S. (in progress). Reading comprehension strategies: Theories, interventions, and technologies. Mahwah, NJ: Erlbaum.

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

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

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

McNamara, D. S., Louwerse, M. M., & Graesser, A. C. (2002). Coh-Metrix: Automated cohesion and coherence scores to predict readability and facilitate comprehension. Unpublished technical report: University of Memphis.

McNamara, D.S. & Scott, J.L. (1999). Training reading strategies. In M. Hahn & S. C. Stoness (Eds.), *Proceedings of the Twenty-first Annual Meeting of the Cognitive Science Society* (pp. 387-392). Hillsdale, NJ: Erlbaum.

Millis, K. K., Kim, H. J., Todaro, S. Magliano, J., Wiemer-Hastings, K., & McNamara, D. S. (2004). Identifying reading strategies using latent semantic analysis: Comparing semantic benchmarks. *Behavior Research Methods, Instruments, & Computers, 36*, 213–221.

Mokhtari, K. & Reichard, C. (2002). Assessing students' metacognitive awareness of reading strategies. *Journal of Educational Psychology, 94*, 249-259.

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

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

Rosenshine, B. & Meister, C. (1994). Cognitive strategy instruction in reading. In D. A. Hayes & S. A. Stahl (Eds), *Instructional Models in Reading*. (pp.85-107). Hillsdale, NJ, England: Lawrence Erlbaum Associates.

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

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

Vygotsky, L.S. (1978). *Mind and society: The development of higher mental processes*. Cambridge, MA: Harvard University Press.

Grant-supported publications (2004 – 2005)

Best, R., Dockrell, J. E., & McNamara, D. S. (2004). Children's semantic representation of a science term. In K. Forbus, D. Gentner, T. Regier (Eds.), Proceedings of the 26th Annual Meeting of the Cognitive Science Society (pp. 1525). Mahwah, NJ: Erlbaum.

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.

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

Graesser, A. C., Hu, X., & McNamara, D.S. (2005). Computerized learning environments that incorporate research in discourse psychology, cognitive science, and computational linguistics. In A. F. Healy (Ed.), *Experimental Cognitive Psychology and its Applications: Festschrift in Honor of Lyle Bourne, Walter Kintsch, and Thomas Landauer* (pp. 183 - 194). Washington, D.C.: American Psychological Association.

Kurby, C. A., Wiemer-Hastings, K., Ganduri, N., Magliano, J. P., Millis, K. K., & McNamara, D. S. (2003). Computerizing Reading Training: Evaluation of a latent semantic analysis space for science text. *Behavior Research Methods, Instruments, and Computers, 35*, 244-250.

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

McNamara, D. S. (2004a). Aprendiendo a través de los textos: efectos de la estructura del texto y las estrategias de los lectores. *Signos, 37*, 123-159.

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

McNamara, D. S., Boonthum, C., Levinstein, I. B., & Millis, K. (in press). Using LSA and word-based measures to assess self-explanationsin iSTART. In T. Landauer, D.S. McNamara, S. Dennis, & W. Kintsch (Eds.), *LSA: A Road to Meaning.* Mahwah, NJ: Erlbaum.

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. *Behavioral Research Methods, Instruments, and Computers, 36*, 222-233.

McNamara, D. S. & O'Reilly, T. (in press). Theories of comprehension skill: Knowledge and strategies versus capacity and suppression. In F. Columbus (Ed.), *Progress in Experimental Psychology Research*. Hauppauge, NY: Nova Science Publishers, Inc.

McNamara, D. S. & Shapiro, A. (in press). Multimedia and hypermedia solutions for promoting metacognitive engagement, coherence, and learning. *Journal of Educational Computing Research.*

Millis, K. K., Kim, H. J., Todaro, S. Magliano, J., Wiemer-Hastings, K., & McNamara, D. S. (2004). Identifying reading strategies using latent semantic analysis: Comparing semantic benchmarks. *Behavior Research Methods, Instruments, & Computers, 36*, 213–221.

Millis, K., Magliano, J., Wiemer-Hastings, K., Todaro, S., & McNamara, D. S. (in press). Assessing comprehension with Latent Semantic Analysis. In T. Landauer, D.S. McNamara, S. Dennis, & W. Kintsch (Eds.), *LSA: A Road to Meaning.* Mahwah, NJ: Erlbaum.

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. (2004a). iSTART: A web-based reading strategy intervention that improves students' science comprehension. In Kinshuk, D. G. Sampson, & P. Isaías (Eds.), *Proceedings of the IADIS International Conference Cognition and Exploratory Learning in Digital Age: CELDA 2004* (pp. 173-180). Lisbon, Portugal: IADIS Press.

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

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.

Todaro, S. A., Magliano, J. P., Millis, K. K., McNamara, D. S., & Kurby, C. C. (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.