C. PROJECT DESCRIPTION

Promoting Active Reading Strategies to Improve Students' Understanding of Science

SPECIFIC AIMS

The problem addressed by this project regards high-school students' understanding of and ability to learn from difficult science texts. The first purpose of this project is to examine high school students' and teachers' knowledge about metacognitive reading strategies to better ascertain the need for strategy interventions, as well as teachers' ability to provide such training to students. The second purpose is to compare the effectiveness of three reading interventions and to thereby identify more effective methods for improving students' comprehension skills. The third purpose is to determine whether the benefits of reading interventions depend on students' skills, knowledge, or cultural environment. The fourth purpose is to develop a computerized training program that adaptively assigns a student to an appropriate training regime, and interactively provides reading training. Hence, this project bridges the two focus areas of the IERI initiative concerning reading skill acquisition and students' understanding of increasingly complex scientific ideas. Within these areas, this project will address issues regarding instructional practices in reading, the role of teacher learning and performance on student learning, optimal interventions for students at-risk of reading or academic failure, the development of better assessment techniques of knowledge and skill, and the use of technology to improve assessment and learning. Importantly, this project addresses these issues at the high-school level. In comparison to students from other countries, it is primarily at this level in our educational system that students in the United States lose ground in terms of their understanding of complex and difficult scientific material (National Assessment of Educational Progress, 1998). Hence, it is imperative to develop interventions that address this problem.

Aim 1. The first aim of this study is to examine high school students' and teachers' knowledge of metacognitive reading strategies. Recent research indicates that metacognitive knowledge is critical to academic performance, particularly metacognitive strategies pertaining to text comprehension. However, there is relatively little research concerning students' knowledge and use of strategies, their impact on academic performance, and how these factors relate to teachers' knowledge of strategy use. Several studies have indicated that teachers' levels of knowledge and use of strategic teaching options in reading and language are far below desirable levels (Conley, 1986; Manzo & Manzo, 1998). This research project will help ascertain how much teachers know about reading strategies and reading training, and how frequently they use such strategies in their classrooms. This information will help us to better understand the need for reading strategy interventions and how teachers can better meet the needs of students.

Aim 2. The second, and principle goal of this study is to compare three reading strategy training programs, Self-Explanation Reading Training (SERT), Previewing, and the Interactive Notation System for Effective Reading and Thinking (INSERT). SERT is based on research demonstrating that when students self explain difficult text aloud, their understanding of the reading material is enhanced (Chi & Bassok, 1989). SERT was designed by the Principal Investigator, Dr. McNamara, to improve undergraduate students' ability to self-explain difficult text by providing instruction and practice using active reading strategies. The SERT program has successfully improved readers' comprehension of difficult science texts and has improved undergraduate students' science examination scores (McNamara & Scott, 1999).

Previewing, or quickly reading over important parts of a text before reading it, has been shown to improve high school students' reading comprehension (Golembesky, Bean, & Goldstein, 1997; Johnston & Allington, 1991; Richardson & Morgan, 2000). Previewing involves skimming a text before reading it and focusing on parts of the text such as key words, headers, the table of contents, summaries, and so on. The goal of previewing is to develop a mental schema of the text's contents before reading it. One underlying assumption is that activating key concepts before reading enables the reader to develop a better understanding of the text. A second purpose of previewing is for the reader to determine the depth at which the text needs to be read given the reader's goals, and appropriate global strategies to read the text (e.g., skimming, rereading, underlining, etc.).

INSERT makes use of a third distinct type of metacognitive operation called marginal glosses (Estes & Vaughan, 1986). A marginal gloss, as outlined by Singer and Donlon (1985) is a method of providing clues in the margins to aid students in their thinking while reading. The INSERT strategy instructs readers to use a notation system to record symbols next to each paragraph to indicate agreement, importance, novelty, and understanding. These notations are intended to cue the reader to think on higher levels of cognition about what is being gained while reading the text.

Like SERT and previewing, INSERT is designed to help students develop metacognitive awareness. However, the three interventions develop metacognition in distinctively different ways. Whereas SERT and INSERT place greater emphasis on local, on-line comprehension monitoring, previewing places a greater emphasis on globally guiding and monitoring comprehension before reading the text. SERT and INSERT differ because SERT provides the student with strategies to repair comprehension problems, whereas INSERT places the emphasis on students' ability to make notes and monitor comprehension while reading. Hence, comparing these three interventions will provide critical insight regarding the relative importance of activity strategy use, schema activation, and comprehension monitoring to effective comprehension.

Aim 3. The third goal is to evaluate how a particular method of teaching metacognitive reading strategies is affected by student characteristics such as prior knowledge, reading skill, or metacognitive awareness. A unique aspect of this study is that the participating students will be from four distinct types of sites: rural (Sumter County, Georgia), rural Appalachian (Floyd County, Kentucky), suburban (Williamsburg-James City County, Virginia), and inner city (Norfolk, Virginia). We will not only examine the three interventions but also determine whether the above mentioned characteristics have any significant impact on student performance. For example, we will determine whether inner city students think differently and perform certain metacognitive operations differently than their rural, suburban, or Appalachian-rural counterparts. We will determine, then, whether there are any significant differences in performance among the four demographic areas.

Aim 4. An important goal of this research program is to develop computer software that administers reading strategy training to students and to subsequently compare the effectiveness of the computerized training program to group training programs. The automated training program will involve video and text presentation of the lecture components, followed by interactive components in which the student types responses into the computer, and the computer program evaluates the responses and provides feedback. The computer will evaluate individuals' responses using recent technology provided by Latent Semantic Analysis (Landauer, Foltz, & Laham, 1998).

Computerized reading interventions will make reading strategy training more readily available to students and potentially adaptable to meet students' individual needs. While developing and examining interventions provides a wealth of information, the benefits of this information are often lost when the support provided by the experimenters and their funding is withdrawn. Too often, the interventions cease or are incorrectly administered after the experimental project has been completed. Developing and testing automated means of providing reading strategy training will increase the probability that the impact of this study will endure after the project has been completed.

BACKGROUND AND SIGNIFICANCE

Metacognition and Reading

Metacognition refers to an individual's monitoring of cognitive processes and the knowledge and use of cognitive processes involved in successful learning (Hacker, Dunlosky, & Graesser, 1998; Reutzel & Cooter, 2000; Tei & Stewart, 1985). A crucial variable in metacognition is the learner's awareness of his or her own characteristics--such as background knowledge, degree of interest, skills and deficiencies--and of the ways in which these characteristics affect learning (Collins, 1994). When applied to reading, metacognition refers to the reader's monitoring of whether the written material is successfully comprehended, coupled with active reading strategies to enhance and repair comprehension. Comprehension monitoring and active reading strategies are increasingly recognized as critical to successful, skilled reading.

Readers better understand and learn more from written material when they monitor their comprehension and use active reading strategies such as previewing, predicting, making inferences, drawing from background knowledge, and summarizing. Skilled readers are more likely to engage in comprehension monitoring and active reading strategies than are less-skilled readers (Brown, 1982; Forget & Morgan, 1997; Long, Oppy, & Seely, 1994; Magliano, Millis, Miller, & Schleich, 1999; Oakhill, 1984; Oakhill & Yuill, 1996). Moreover, providing readers with instruction to use metacognitive reading strategies improves reading comprehension skills (Baker, 1996; Baumann, Seifert-Kessell, & Jones, 1992; Davey, 1983; Dewitz, Carr, & Patberg, 1987; Hansen & Pearson, 1983; McNamara & Scott, 1999; Palinscar & Brown, 1984; Yuill & Oakhill, 1988). Surprisingly, teachers often have little knowledge of the importance of metacognition in the reading process (Conley, 1986; Manzo & Manzo, 1998). Moreover, there is relatively little research concerning high school students' knowledge of and use of strategies, its impact on academic performance, and how that relates to their teachers' metacognitive knowledge and classroom instruction. We propose that a better understanding of these relationships will help us to improve educational practice. Our contention is that metacognitive awareness on the part of students can be improved through classroom instruction whereby students are taught metacognitive strategies as they learn the content of the material. This research will help ascertain teachers' knowledge of reading strategies and reading training, and how frequently they use such strategies in their classrooms. This information will help us to better understand the need for reading strategy interventions and how teachers can better meet the needs of students.

Reading Interventions

Textbooks are the source of a large part of what students learn during their academic career. Hence, reading is a critical skill for a student to perform well academically. However, many students are poor readers, or have difficulty understanding expository texts (Bowen, 1999). Nevertheless, recent research has shown that providing students with reading strategy instruction and training improves reading comprehension. This section reviews literature regarding the three reading strategy interventions targeted in this project, self-explanation, previewing, and INSERT. These three interventions are designed to improve comprehension of difficult text for readers who have sufficient word decoding skills, but have difficulty comprehending written material. Research indicates that there are many poor comprehenders who do not have deficits at the level of word decoding or syntactic decomposition. These readers will read fluently and with apparent ease, and yet still perform poorly when answering questions about text (e.g., Stothard & Hulme, 1996). The self-explanation, previewing, and INSERT interventions target processes believed to be important to comprehension skill and comprehension monitoring, but differ in terms of which processes are emphasized. Therefore, comparing the three interventions will indicate the relative importance of the processes emphasized by each technique for improving readers' comprehension of difficult text.

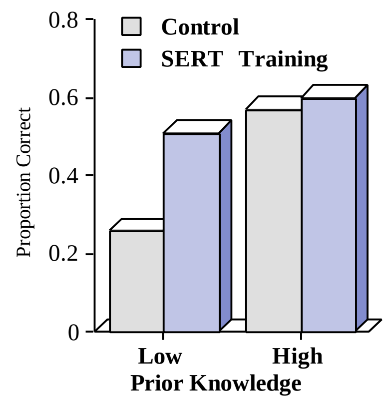
Self-Explanation Reading Training (SERT)

Self explanation refers to the process of explaining text while reading. This process requires actively processing the text, understanding the relationships between separate ideas in the text, and relating the ideas to knowledge already possessed by the reader. In sum, a skilled reader reads text to understand it. In a laboratory setting, self-explanation involves reading and explaining aloud sentences or sections from a text. Readers who explain the text either spontaneously, or when prompted to do so, understand more from the text and construct better mental models of the content (Chi & Bassok, 1989; Chi, de Leeuw, Chiu, & LaVancher, 1994; Magliano, Trabasso, & Graesser, 1999; Trabasso & Magliano, 1996). However, some readers do not naturally self-explain text, and self-explain poorly when prompted to do so. Therefore, SERT was designed by the Principal Investigator, Dr. McNamara, to improve college students' ability to self-explain difficult text by providing instruction and practice using reading strategies (McNamara & Scott, 1999).

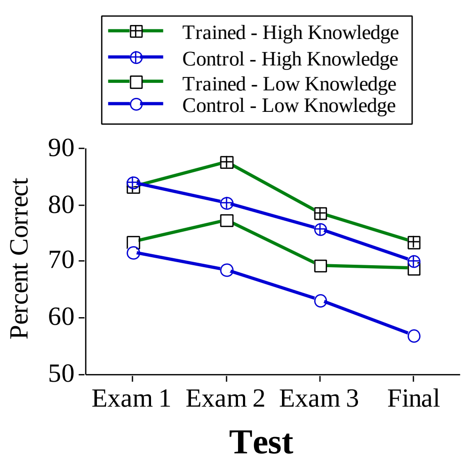
SERT was inspired by previous research showing the benefits of strategy instruction (Bielaczyc, Pirolli, & Brown, 1995; Palinscar & Brown, 1984; Yuill & Oakhill, 1988). The SERT intervention is much like techniques based on thinking aloud (Baumann, Seifert-Kessell, & Jones, 1992; Davey, 1983; Coté, Goldman, & Saul, 1998). However, SERT places greater emphasis on use of active reading strategies to explain the text than have previous thinking aloud interventions.

SERT training can be administered to a small group of students and requires at least 2 hours to administer. It begins with a brief instruction including definitions and examples of self-explanation and reading strategies. The strategies focus on the benefits of predicting what the text will say, making bridging inferences between separate ideas in the text, using prior knowledge and logic to understand the text, and monitoring comprehension. After this brief instruction, students then read a science text and watch a video of a student in the process of self-explaining the text. The students identify at certain points in the text the strategies used by the student in the video, and then discuss these strategies as a group. The students then work with partners to practice the strategy, taking turns in reading orally and sharing thoughts. Instructors are present to assist and monitor the students.

One particular goal of the SERT research has been to determine whether and how readers' prior domain knowledge of the text interacts with the effectiveness of self explanation. Research has indicated that knowledge about a text domain is particularly important to understanding (McNamara & Kintsch, 1996). One question has been whether SERT training would help low-knowledge readers to overcome their barriers. The figure to the right shows comprehension accuracy for high and low-knowledge readers after self-explaining a science text (McNamara & Scott, 1999). Those who had received training used more effective strategies and better understood the text. Moreover, SERT was most effective for students who had less prior knowledge about the text domain. This training provided students with strategies that they could use while reading that effectively compensated for their lack of domain knowledge. In addition, protocol analyses indicated that these readers instead relied on their common sense and logic to understand the text.



Three subsequent experiments have shown that SERT training not only improves text comprehension, but also improves undergraduate students' exam performance in science courses. Across five classrooms including almost 1000 students, consistent benefits have been found for SERT (McNamara, 2000). Reliable advantages on exam scores for students who received SERT training in comparison to Control students have ranged between 5 and 14 percent. In addition, across all five classrooms, prior knowledge of scientific facts showed the strongest correlations with exam performance, whereas prior reading skill showed the lowest correlations (which were generally non-significant). Most importantly, training generally had the greatest benefits for those students with less prior knowledge about science. For example, the figure to the



For example, the figure to the left displays the results from an Ecology course including 92 students, 33 of whom had participated in SERT. Training was administered between Exams 1 and 2, resulting in a substantial improvement on Exam 2 for those who received training in comparison to control students. In addition, there was a decline in performance across exams for all students except low-knowledge students who received SERT training. Indeed, these low-knowledge students showed comparable performance on the final exam to high-knowledge students in the control condition who did not participate in SERT.

Within all of these experiments, low-knowledge students who did not receive training often left the science course without a passing grade. Hence, for some students, SERT translated to a difference of passing and not passing the course. Nevertheless, many of the high-knowledge students, who had more knowledge about the targeted course, reported that they had used the strategies for more challenging courses in which they were enrolled--and that these strategies were effective. Therefore, this type of training has the potential to be beneficial to many students, regardless of their prior knowledge.

Measures of reading skill were unable to predict who benefited from reading strategy training. It may seem counterintuitive that a standardized reading measure could not predict who will benefit most from training--after all the training is geared toward improving reading ability. However, the reading ability that is captured by reading skill measures is not equivalent to a student's ability to use reading strategies that improve learning from text. One important goal will be to determine whether a measure of strategic reading is better able to predict the utility of this training program, regardless of the student's knowledge level.

Previewing

Previewing, a process that is important for clarifying the cognitive structure of students before reading, has been found to be effective in improving reading comprehension (Golembesky et al, 1997; Short & Ryan, 1984). In the previewing stage, students first skim key sections of the material for the purpose of selecting strategies appropriate to the depth and duration of study needed. For example, when previewing a technical chapter or a report, students are taught to examine and think about the following: (1) Title and subtitle—to discover the overall topic of the chapter or article; (2) Author's name—to see whether the author is a recognized authority; (3) Copyright—to see whether the material is current; (4) Introduction—to learn what the author intends to talk about; (5) Headings and subheadings—to identify the topic of the sections that follow; (6) Graphs, charts, maps, tables, pictures—to aid in understanding specific aspects of the material; (7) Summary—to get an overview of the reading; and (8) Questions—to review important topics covered in the chapter. To select proper strategies—whether note taking, underlining, or rapid reading—students must spend time clarifying their thinking about the topic. Then they need to ask themselves questions such as the following: How interested am I in this section?; How deeply do I need to think and concentrate to learn this material?; How fast can I read this material?; and What do I need to learn about this topic?

Previewing reduces uncertainty about a reading assignment, allowing students to gain confidence, read in a more relaxed manner, gain interest, and improve their attitude toward the material. In addition, previewing strategies enable students to decide how much of the material is in their own background of experience. As a result of a previewing strategy, learners are clearer about what they know and what they need to know. In effect, they set a purpose for reading before they begin.

When providing previewing instruction, the teacher assists students in deciding what they already know about the material and what they need to learn. The reader turns those things that are now known into questions, which provide a purpose for reading. Students reading fiction need to preview the title, illustrations, and introduction in order to make hypotheses about the outcome of the story. This preview heightens suspense and aids in maintaining interest. Most important, predicting of study structure gives the reader a purpose for reading—namely, to find out whether the predictions are correct. Whether students are reading fiction or expository or informational material, a very important reason for previewing is that it forces them to do the sophisticated kind of thinking required for drawing inferences and developing interpretations. Students generally will not preview on their own unless teachers model and provide practice in this skill (Richardson & Morgan, 2000).

Interactive Notatation System for Effective Reading and Thinking (INSERT)

The INSERT method, developed by Vaughan and Estes (1986), is a strategy that students can use to monitor their own comprehension as they read. This technique employs a marginal glossing system similar to that first developed by Singer and Donlon (1985), whereby the reader records a set of symbols by each paragraph to monitor agreement or disagreement with the reading, whether the reader thinks the reading is important, whether the reading is new to the reader, and whether or not the reader understands what is being read. This system of comprehension monitoring is designed to aid students in becoming aware of their own thinking (Vaughan & Estes, 1986). To our knowledge, there have not been any empirical tests of the effectiveness of INSERT. Hence, this study will provide valuable evidence in this regard.

Comparing the three interventions

All three of these interventions are designed to improve readers' ability to monitor their comprehension. However, they differ in terms of the targeted cognitive processes. Specifically, SERT emphasizes the use of reading strategies to explain and repair comprehension problems while reading; previewing emphasizes schema activation and global reading strategies before reading; and INSERT emphasizes note taking as an overt means of monitoring comprehension while reading. Hence, comparing these three interventions will provide critical insight as to the relative importance of active strategy use, schema activation, and comprehension monitoring to effective comprehension. As found previously, we hypothesize that SERT will be most effective for students who possess less knowledge about the target domain. SERT is also predicted to be effective for students with less knowledge about metacognitive reading strategies. In contrast, previewing is expected to be most effective for high-knowledge students. That is, we hypothesize that previewing will primarily aid students who have the prior knowledge necessary to understand the concepts encountered while previewing. We do not expect INSERT to be effective in comparison to either SERT or previewing because it does not target cognitive processes essential for repairing comprehension problems. Moreover, it is an overt action (taking notes) that may distract from the reading process, and may not generalize to all reading situations. Hence, INSERT is intended to serve as a control, or baseline condition.

Computer-based Evaluation and Tutor of Effective Reading Strategies

A critical goal is to develop a computerized tutoring program that implements the most successful of the three training approaches. If more than one approach successfully improves comprehension and academic performance in comparison to a control condition, then the computer-based tutor will include aspects from each. Because of its great initial success, we expect that the tutoring program will contain aspects of SERT.

We envision that the tutoring program will consist of three major components: Assessment, Strategy training, and Practice. The first component will assess users' needs for training. The particular methods of assessment will be guided by the results of our experiments, but may include measures of reading skill, metacognitive strategy knowledge, or prior domain knowledge. The assessment will also include an evaluation of students' verbal protocols to specific sentences embedded in expository texts. That is, the students will read texts one sentence at a time, and after some sentences, will be asked to enter their thoughts. This component will identify strategies that users employ without prior training. This data will serve as a baseline measure of each user so that the system can determine whether there has been an increase in effective strategy use during and after training. The second component will illustrate and explain the key reading strategies, as identified by the earlier described experiments. This information will be presented via multimedia, including text and video presentations. The third component will allow the user to practice the strategies while reading texts. The computer will evaluate the responses and provide fundamental feedback and hints when appropriate. For example, the computer might respond with "Try thinking back to other sentences" if the computer determines that the user tends to only paraphrase the current sentence. All components will be available to the user at any given time. This would allow the user to retake the assessment component after going through the tutoring component, so that the computer could note improvement, passing that information on to the user.

To implement the computer-based tutoring program, it is necessary for the computer to 'understand' and evaluate written language and discourse. The complexity of human discourse has impeded past efforts to develop such programs. One barrier has stemmed from human utterances that are frequently ill formed syntactically or semantically. A second impediment has been the large body of knowledge necessary for a computer to understand human language. The proposed tutor will overcome these barriers with the help of Latent Semantic Analysis (LSA; Landauer & Dumais, 1997). LSA is a text-processing tool that represents the semantic contents of text units based on their co-occurrence frequency with all other text units within a large corpus of text. First, LSA computes a matrix of how frequently individual words co-occur with each other within all documents of text in the database. In the proposed application, this text database would consist of an extensive set of verbal protocols within a scientific domain, e.g., cell mitosis. Scientific texts would enlarge the text corpus, resulting in a large database that would produce meaningful co-occurrence matrix. The matrix is then transformed: an algorithm called Singular Value Decomposition (SVD) is applied to the matrix to reduce the dimensionality to an "ideal" number. This number is determined empirically by assessing how well LSA text evaluations match the evaluations of domain experts. In the resulting high-dimensional semantic space, each text unit is represented as a vector with as many elements as there are dimensions. When presented with two text units, LSA computes their similarity by computing the cosines of their vectors. The cosine measures the similarity of the two vectors across all dimensions. The more similar the vectors are, the higher the LSA cosine is. Cosines of 1.0 indicate high similarity. The minimal cosine is 0.0, and indicates maximal dissimilarity. LSA cosines of text units, both words and paragraphs, have been shown to reliably match human similarity judgments of documents (Landauer & Dumais, 1997; Landauer, Foltz, & Laham, 1998; Landauer, Laham, Rehder, & Schreiner, 1997).

LSA will provide the critical tools necessary for the automated tutor to 'understand' a user's inputs. Understanding will be achieved by assessing the semantic similarity between the user's response and some semantic benchmark – a body of text that can be compared to the user's response. The availability of verbal protocols from previous research will facilitate the development of such benchmarks. Previous research has shown that matching user inputs and benchmark documents with LSA yields results that can be used reliably to assess the user's ability. For example, Wolfe and colleagues (1998) have shown that LSA can be successfully used to assess readers' knowledge in a domain by comparing their essays to texts of varying difficulty. Foltz, Laham and Landauer (1999) have successfully used LSA to grade essays based on the match between the new essay with a set of graded essays. Finally, Graesser and his colleagues have implemented LSA in AutoTutor, a computerized tutor that evaluates student contributions based on LSA matches with ideal versus bad answers (Graesser, Wiemer-Hastings, Wiemer-Hastings, Harter, Person & TRG, in press; Graesser, Wiemer-Hastings, Wiemer-Hastings, Kreuz, & TRG, in press; Wiemer-Hastings, Wiemer-Hastings, & Graesser, 1999).

One semantic benchmark that we will consider is responses produced by previously identified expert readers. For example, let us consider a user who types in "the blood needs to go to the left ventricle to get oxygen" after reading the sentence "The blood enters the left chamber" which is embedded in a passage about the heart. LSA can compute a cosine between the user's response to the sentence and stored responses by experts (e.g., "blood needs oxygen," "afterwards the blood travels to the left ventricle," "it is pumped to the left ventricle," etc.) In this case, a high cosine would occur, indicating on-line understanding of the text.

LSA and the general approach outlined above provide an architecture that is flexible, yet powerful enough to implement various forms of the computerized tool. Flexibility is needed because it is unclear at the start which strategy will be most effective. As mentioned earlier, however, it is likely that the tool will implement the main ingredients of the SERT approach. In order for the tutor to implement these strategies, the computer will need to recognize paraphrases, bridging inferences, elaborations, and predictions. The words that will comprise the "semantic benchmark" for each of these categories will be taken from the verbal protocols collected from skilled readers in Experiments 2, 3, and 5, as well as protocols previously collected by McNamara and Scott (1999). Therefore, there will be a set of paraphrases (PP), bridging inferences (B), elaborations (E), and predictions (P) associated with each sentence. The tutor will compute a cosine between the input and each of these benchmarks. The pattern of cosines across these categories will reveal the statement type of the user. For example, the pattern PP > (B, E, P) indicates a paraphrase. If the cosines across the categories are equally low, then this would indicate that the user is not saying anything relevant. If equally high, the pattern would then not indicate a particular statement type. To help limit this possibility, care will be taken to choose target sentences (i.e., sentences that the user will respond to), which resulted in different words (from the protocols) across the categories.

The magnitude, consistency, and pattern of the cosines across the categories will trigger different responses from the tutor that will help shape the student's use of the strategies. For example, consider a student who inputs a prediction early on in the session. The computer might respond by "Good, you are thinking about what might come up." However, if this pattern is consistent across sentences, then the computer might respond with "You might try considering how the current sentence relates to previous sentences." The computer might also give an example of a bridging inference. Consequently, if a person appears to be consistently paraphrasing, then the computer will suggest that he or she predict. When there is no decipherable pattern, the computer will either not respond or respond with a neutral "OK." The specific responses will be guided by previously collected recordings of human tutors administering self-explanation training.

Recent work by Magliano and Millis (2000) provides a basis for the proposed analysis and suggests that LSA can be used to evaluate verbal protocols. Magliano and Millis had readers think aloud to sentences of a simple story. They then used LSA to estimate the semantic similarity between readers' think aloud responses and specific sentences in the story. A causal overlap proportion (COP) was calculated based on the average cosine between the response and causally related sentences in the text (C), and the current sentence (CS). The COP score was defined as C/(CS+C). Hence, a higher COP score indicated greater effort toward establishing explanatory coherence as opposed to paraphrasing the current sentence. They found that the COP scores significantly predicted scores on the Nelson Denny test of reading comprehension and recall. Therefore, the COP scores correlate with measures of comprehension. The results of this study suggest that LSA can be used to compare verbal protocols to semantic benchmarks and that the results of that comparison are indicative of comprehension.

PROPOSED WORK

This section describes our proposed reading interventions, skill and knowledge measures, the experiments, and the development of the automated tutoring system. Our immediate goals are to establish relationships with science teachers in the participating schools and to identify teachers to participate in this project. It is essential that the results of this study generalize to a diverse population. Hence, as we stated earlier, we will include students from four distinct sites within three states. Specifically, the public high school officials in Norfolk, VA; Williamsburg-James City County in Williamsburg, Virginia; Floyd County Schools in Prestonsburg, Kentucky; and Sumter County High School in Americus, Georgia have committed to participate in this project. This diversity insures that we have included students from a variety of ethnic and socioeconomic backgrounds.

Reading interventions

Self-explanation reading training (SERT)

The current SERT program consists of three phases: instruction, strategy identification, and practice. Students are first given a half-hour scripted instruction in self-explanation and reading strategies, focusing on the benefits of using prior knowledge and logic to understand the text, predicting what the text might say, making bridging inferences between separate ideas in the text, and monitoring comprehension. Students follow the lecture with a handout that describes and provides examples of each strategy, while asking questions for clarification during the lecture. Students then identify strategies exemplified within a video presentation of another student self-explaining a science text. The training video was created by compiling skilled self-explanations of the text and then creating a script for a student to act out the process of self-explaining the text. The instructor stops at four predetermined points in the video and asks the students to identify the strategies used by the student in the video. After all of the students have done so, the instructor leads a discussion concerning which strategies were used. The students then practice in pairs by taking turns self-explaining each paragraph of two texts. After one student has self-explained each sentence of an entire paragraph, the second student summarizes the paragraph, and then self-explains the following paragraph, and so on. The summarization procedure ensures that the "listening" student is attentive while the other student self-explains. In addition, the process of summarization enhances text understanding.

SERT will be modified to meet the needs of high-school students. First, the practice texts will be changed to science texts that are at the appropriate level. The texts to be used have been chosen. However, pilot experiments with paid high-school participants will be conducted to ensure that these texts are appropriately challenging. Second, two additional 1-hour sessions will be added to follow-up training. These sessions will be added due to observations that the effects of training have tended to wane as a function of time. The follow-up sessions are expected to mitigate this forgetting effect. The follow-up sessions will consist of a short review and discussion of students' issues, comments, and problems concerning the strategy, followed by additional practice using the strategy.

Previewing training (PRE)

Previewing training will consist of a brief instruction concerning the concept of previewing, examples of sections of text that should be previewed, and the importance determining reading needs and goals to select appropriate reading strategies. The instructor will provide examples of texts for which key sections to preview are pointed out, along with examples of situations that would affect a reader's choice of strategies. This is a silent process, and thus a video presentation would not be appropriate in this case. The students will then preview and read two texts. Two follow-up sessions will be provided to discuss and practice the previewing process.

Interactive notatation system for effective reading and thinking training (INSERT)

Students will be trained in how to use marginal glosses in general and in the INSERT technique in particular. Instructors will demonstrate in this phase how the INSERT works by taking students through a reading and using the notation system to indicate how the teacher is monitoring his or her own comprehension. After such guided practice, students will use INSERT over two readings and two follow-up sessions will be provided to discuss and practice the technique.

Measures

Reading Skill

Reading skill will be measured with the Gates-MacGinitie Reading Test for grades 10-12 (Forms K and L) which has grade norms available for grades 7.0 to 12.9. This test is widely used and has employed rigorous standardization testing. The comprehension test involves reading short passages and answering multiple-choice questions. The vocabulary test involves choosing the correct meaning of a word among four alternatives. Reading skill will also be measured in two other ways. First, available state tests of achievement in reading and reading-related areas (e.g., Standards of Learning Tests) administered by the states in Georgia, Virginia, and Kentucky will be utilized. Also used will be a cloze test on a reading passage from books used in the intervention sessions. The passages will be similar to ones used in the intervention.

Metacognitive Knowledge

Metacognition will be measured with the Metacognitive Strategies Index (MSI), first developed by Schmitt (1990). The MSI is multiple choice questionnaire consisting of 25 questions which measure six factors associated with metacognitive reading strategy knowledge and use: (1) predicting and verifying, (2) previewing, (3) purpose setting, (4) self-questioning, (5) drawing from background knowledge, and (6) summarizing. Forget (Morgan, Forget, & Antinarella, 1996) adapted Schmitt's elementary test for secondary populations. Forget found validity and test-retest reliability to be fairly high in content area studies he conducted at the high school level (Forget, 1991; Forget & Morgan, 1997; Forget, 1999).

Course Domain Knowledge

Course domain knowledge refers to the student's prior knowledge on which a course builds; essentially the prerequisite knowledge that should have been acquired during a previous course. A course domain knowledge measure will be developed for each targeted course topic. To do so, multiple choice questions will be developed based on the information covered within the textbook or course that was a prerequisite to the target course. The measure will include factual, conceptual, and applied questions. Pilot experiments will be conducted in Years 1 and 2 to establish the reliability and validity of the items and measures.

Text Domain Knowledge

Text domain knowledge refers to prior knowledge that is necessary to understand a text. Text domain knowledge measures will be developed using the same techniques as described above for course domain knowledge. Pilot experiments will be conducted with paid high-school students to ensure the reliability and validity of each prior knowledge measure.

Text Comprehension

Text comprehension measures are developed based on models of text comprehension (Kintsch, 1988). Specifically, we assume that there are various levels of comprehension, and that different types of questions assess these levels of understanding. The guidelines for developing questions are to include questions that target single ideas (text-based), related ideas that are separated in the text (bridging inference), conceptual questions that target understanding of the text as a whole, and problem solving questions that require applying the information to novel situations (McNamara, Kintsch, Songer, & Kintsch, 1996). Pilot experiments will be conducted with paid high-school students to ensure the reliability and validity of each comprehension measure.

Experiments

Question 1; Metacognitive Knowledge

Our first question deals with how much high school students and their science teachers know about metacognitive and reading strategies. This research will ascertain students' metacognitive knowledge and its relation to academic skills and performance. This research will also further investigate teachers' knowledge of metacognition and whether that translates into better teaching concerning appropriate use of reading strategies with text.

Experiment 1. Our first experiment will be a correlational study of the relationships between students' and science teachers' metacognitive knowledge, teachers' use of strategies in the classroom, students' knowledge about the course domain, and students' academic performance. In addition, we will examine whether there are differences between these variables as a function of site. As mentioned earlier, our target sites differ in terms of the students' academic and socioeconomic environments, including rural, rural Appalachian, suburban, and inner city school systems. This study will provide valuable information concerning differences between teachers and students as a function of these environments. This study will also provide baseline data for the subsequent experiments.

Students and teachers from 10 to 15 science classrooms from each site will be included this study. Students and teachers will be administered the metacognitive strategies index at the beginning and end of the school terms (i.e., pretest, posttest). Students will also complete the course domain knowledge measure at the beginning and end of the school terms. Teachers will be observed and asked to complete a questionnaire to assess their use of strategies in the classroom. The questionnaire will be modified from one developed by Forget (1999). The results from these measures will be correlated with student outcome measures including course performance and standardized measures of learning (i.e., school-administered SOL tests). Differences between sites will be examined using multiple regression. We expect that (a) students whose teachers use metacognitive strategies in the classroom will have higher metacognitive strategy knowledge; (b) strategy use will be highly correlated with students' academic performance and SOL measures; (c) correlations between metacognitive knowledge and student performance will be stronger for students who have less prior knowledge about the course domain; and (d) students and teachers from suburban schools will be more likely to use metacognitive reading strategies.

Questions 2 and 3: Reading Intervention Effectiveness

The second question to be addressed by this research project is whether the three training programs differ in terms of their impact on students' ability to learn from difficult science texts at the high school level. This impact will be measured in terms of the students' comprehension of science texts, performance on course examinations, and overall academic performance. The third question is whether the effectiveness of a certain training program depends on learners' knowledge of strategies and their domain knowledge of the material to be learned. To this end, we will assess students' reading skill, metacognitive knowledge, and prior knowledge of the text and/or course domain.

Experiment 2. The second experiment will serve as a pilot study to test and refine the three training conditions. The experiment will include 72 high-school students (who have obtained parental consent) from Norfolk High School (i.e., at least 24 per training condition) who will be randomly assigned to one of three experimental conditions (SERT, PRE, INSERT) and one of two trainers (Trainer A, Trainer B). The experiment will include three sessions. Pretest measures of reading skill and metacognitive knowledge will be administered in the first session. Results from the reading skill measure will be used to screen students with severe decoding deficits (those students will be recommended for remedial reading). The pretest measures will also be used to match students across the three conditions. Training will consist of one 2-hour session administered to approximately 12 participants at once (all training sessions throughout this study will be recorded). Two training groups will be conducted for each condition; each group being trained by a different experimenter. A third session (within 1 week of training) will be administered to each student individually and include comprehension tests for two science texts, each from different domains (with order of text presentation counterbalanced across conditions). The students will be asked to use the techniques that were taught during training while reading the texts. This session will be recorded such that SERT participants' self-explanations can be transcribed and analyzed. After completing the comprehension test, participants will be asked via a retrospective questionnaire what strategies they had used while reading the texts. The questionnaire is necessary to ascertain the use of prior reading strategies, as well as strategies learned from students in other conditions (i.e., treatment diffusion).

Separate analyses of covariance will be conducted on comprehension scores for each text including the experimental factors of condition and trainer, with students' scores on the skill and knowledge measures as covariates. These results will provide preliminary information concerning the relative effectiveness of the training programs. Differences as a function of the two trainers would indicate a need to further refine the training protocols. Students' and trainers' qualitative judgements of the training programs will be also be assessed to determine whether the training programs need to be revised to better fit the needs of high-school students. If the training regimes require substantial modifications, this experiment will be repeated with a separate pool of students.

Experiment 3. The third experiment will use similar methodology to Experiment 2, but will be a more extensive study to compare the effectiveness of the three training conditions. This study will also examine effects of training on exam and course performance, as well as potential interactions of training condition with student' skills, knowledge, and cultural environment. The experiment will include 1200 high-school students (who have obtained parental consent) enrolled in science courses across the four sites noted earlier (i.e., approximately 100 students in each experimental condition at each site). The large number of participants is required to examine possible differences between training sites, to examine potential interactions of condition with individual differences (particularly domain knowledge and metacognitive knowledge), and to accommodate for possible attrition. A pretest session administered in large groups will include measures of reading ability, metacognitive knowledge, and course knowledge. Students from participating classrooms will be matched according to their abilities and then randomly assigned to one of three experimental conditions (SERT, PRE, INSERT) with approximately equal numbers of students from each classroom across conditions. As in Experiment 2, training will consist of one 2-hour session administered to groups of approximately 12 students. The experiment will also add two refresher-training sessions. (Note: If the results of Experiment 2 indicate a need, a control condition will be added to this study who will read the same texts as do the training conditions, but will not be provided with any strategy instruction.)

A third session (within one week following the primary training session) will be administered to assess comprehension of two science texts. The students will be asked to use the techniques that were taught during training while reading the texts. Because of the large number of students, it is anticipated that individual sessions will not be feasible for all students within a reasonable time period. Hence, this session will be administered in groups for the majority of the students (students who had received SERT training will be asked to self-explain quietly or silently). However, a subset of students will be administered testing individually. Specifically, 60 students (20 from each training condition) will be identified from the lower, middle, and upper tiers in terms of knowledge and skill. These students will all be asked to "think aloud" while reading the texts. Participants' utterances will be transcribed and scored by classifying the utterance according to strategy type. Utterances will also be scored using LSA analytic tools. For example, we will compare whether the utterance has a higher cosine with the focal sentence (indicating a paraphrase) or with previous sentences in the text (indicative of a bridging inference). These results will provide evidence that LSA can distinguish between utterance types, and that LSA protocol scores are predictive of text comprehension scores.

A fourth and a fifth session will be administered to provide refresher training. During these sessions, the strategies will be reviewed and discussed. The students will then read two texts using the appropriate strategies and answer questions about the texts. Students in the SERT training condition will read aloud the texts with partners. These interactions will be recorded.

After each exam, students will be asked what strategies they had used to read the material covered by the exam, and how they studied for the exam. In addition, a final questionnaire will be administered at the end of the course to ascertain what strategies the student used during the course, and whether the strategies they had learned were used for other courses in which they were enrolled during that year. This questionnaire will also collect qualitative judgments of training effectiveness. Similar questionnaires will be administered to teacher participants.

An important question addressed here is whether the effectiveness of training depends on students' individual differences. The two variables of greatest interest are prior domain knowledge and metacognitive knowledge. Prior knowledge is important because it provides a schema to better understand the information, and when the information is linked with prior knowledge it is better retained. Metacognitive knowledge is important because reading strategies can help the reader to overcome knowledge deficits. We hypothesize that SERT will be most effective for low knowledge students, and for students with low metacognitive knowledge about reading strategies. In contrast, PRE will be most effective for high-knowledge students. That is, we hypothesize that previewing will primarily aid students who have the prior knowledge necessary to understand the concepts encountered while previewing. We hypothesize that the INSERT training will not significantly aid students.

There are two common statistical methods for examining individual differences within quasi-experimental studies. The first is to include as a covariate participants' scores on the variable of interest (as in Experiment 1). However, analysis of covariance removes the variance associated with that variable from the analysis, and does not allow examination of possible interactions between the measures of covariance and the experimental variable. The second method is to categorize participants as high or low ability according to a their score on the measure(s). This method produces relatively clear results, but loses the power afforded by the continuous variable, and potentially ignores variance within each category (e.g., curvilinear effects). A third alternative, which will be adopted to analyze the results of Experiment 3, is to conduct the analyses of variance using multiple regression (e.g., Judd & McClelland, 1989). This approach affords looking at effects of experimental factors and continuous measures of individual difference simultaneously, thus revealing potential interactions. In this way, analyses will be conducted to examine potential interactions between training condition, site, and individual difference measures. Dependent measures will include text comprehension, exam scores, and overall course performance. Students' performance in concurrent courses will also be examined to measure how well training effects generalize across domains.

Experiment 4. Experiments 2 and 3 will indicate which of the three training approaches is most effective, and whether individual differences or site differences interact with training effectiveness. Experiment 4 will examine the effectiveness of integrating reading strategy training within science courses. The classrooms of interest will be comprised of predominantly at-risk students in the 9^{th} and 10^{th} grades. Thus, the intervention will target the reading strategy intervention most effective for at-risk students. If the interventions were found to be equally effective, SERT training will be combined with PRE and INSERT training. The classroom teachers, who will be trained before the start of academic term (i.e., during the month of August), will administer the reading intervention throughout the term in the context of the classroom text materials. Teachers' sessions will be recorded and teachers will be asked to complete weekly activity logs. The dependent measures will include students' exam and course performance. Comparisons will be made to classrooms that are also predominantly comprised of at-risk students. Within-subject comparisons will also be made to students' performance in their previous courses and other current courses (converted to Z-scores). Baseline performance measures for these courses will be collected during Years 1 and 2. Measures of covariance will include students' prior knowledge of the course content area and school-administered achievement test scores. The experiment will include approximately eight classrooms identified from each of four training sites. Half of the teachers will be randomly assigned to receive training, and the remaining will serve as control classroom teachers. Only teachers will be included in this study who do not normally instruct their students about reading strategies. Class size and class demographics will also be considered as having possible effects on training effectiveness; an attempt will be made to use classes of relatively equal size.

Question 4: Automated Reading Strategy Instruction

This question deals with how to make reading strategy training widely and easily available to students. Our answer to that question is to develop a computerized training program. This training program will involve online assessment, a video and text presentation of the lecture component, followed by interactive components in which the student types responses into the computer, and the computer program evaluates the responses and provides feedback. The computer will evaluate responses using recent technology provided by Latent Semantic Analysis (LSA). We expect that the development of this program will require three years. Much of this effort will involve computer programming. It will also involve a great deal of protocol analyses, and iterative examinations of LSA's ability to predict students' comprehension. During Years 3 and 4, preliminary experiments will be conducted with paid participants to examine the usability and apparent effectiveness of the system. We expect that the system will require several iterations before testing it within the high schools. We are confident that the system will be completed by Year 4. At that time, we will conduct Experiments 6 and 7 to examine the relative effectiveness of our automated tutoring system.

Development of the automated tutor. This process involves three major programming objectives. The first objective is to automate the presentation of the reading strategy instruction. The second objective is to develop the interface between the instructional and LSA components of the computer program. The third objective is to build the necessary LSA components. Much of this latter process was described earlier in the section *Computer-based Evaluation and Tutor of Effective Reading Strategies.* Developing the LSA components involves the following subgoals. First, the LSA system will be trained on the domains relevant to the training texts. Second, appropriate tutorial responses by LSA will be determined. Third, the cosines between verbal input and LSA will be determined. Fourth, these cosines will be coordinated with tutorial responses. For much of this work, the necessary verbal protocol data has already been collected. We have available 70 readers' verbal self-explanations during and after training, along with their human tutors' prompts during training. This data will help us to compute LSA analyses of readers' explanations of text, and the relationship between these analyses and readers' skills. Nevertheless, it is anticipated that we will need additional data to achieve our goals. Experiment 5 is designed to provide such data.

Experiment 5. Students will be provided with SERT training in groups. The students will then participate in individual sessions to self explain the texts that are to be used within the automated tutor. These verbal protocols will provide additional comparisons for LSA to compute the quality of particular utterances. In addition, correlations between LSA analyses of these verbal protocols and readers' skills will further indicate the success of LSA in predicting reading ability and metacognitive knowledge. This experiment will be repeated if necessary.

Experiments 6 and 7. The purpose of these experiments is to examine the effectiveness of our automated tutor. In Experiment 6, we will compare the benefits of the automated tutor to a control condition and to group training sessions. In Experiment 7, we will compare the automated tutor to a control condition and to the integration of strategy training into the classroom. The methodology will be similar to those employed in Experiments 3 and 4. However, as in Experiment 4, classrooms will be randomly assigned to condition. Each classroom assigned to receive automated tutoring sessions will be provided with one or two computers (as deemed necessary) if appropriate computers are not readily available within the high school. These experiments will indicate whether the automated tutor effectively improves students' reading and learning skills in comparison to no training, and whether the tutor is as effective as training administered by humans.

QUALIFICATIONS OF PROJECT INVESTIGATORS

We are in a unique situation and have unique qualifications that will allow us to complete the proposed work as an interdisciplinary team. Danielle McNamara, a cognitive scientist, has conducted research on text comprehension for the past 10 years. Her earlier work addressed issues concerning the effects of text structure and familiarity on readers' ability to understand and learn from expository texts. For the past 5 years, she conducted research regarding the effects of strategy instruction on readers' text comprehension and course exam performance. She has also conducted research to investigate the use of LSA to predict readers' comprehension of texts. Although she is a relatively young investigator, she has previously served as PI on five grants. Hence, she has had experience managing and conducting large-scale projects of this nature. Ray Morgan has had over thirty years of experience in reading education and conducting educational research. He currently serves as a reading education consultant for numerous schools across the United States. His experience with high school students and high school teachers in particular offers unique qualifications that will undoubtedly enhance the quality of this study, as well as the likelihood of its success. Irwin Levinstein is a computer scientist who has been an active participant in several Computer Science Department curricular initiatives. In particular, he has worked on the O.D.U. Teched initiative to improve reading skills of students who are seeking bachelor's degrees from remote locations over the internet. Hence, his prior experience in developing reading tutorials is a cornerstone to this project.

The investigators at Northern Illinois University bring to this project their expertise in reading comprehension, reading skill, latent semantic analysis, and automated tutoring. Keith Millis has done extensive research on the role of individual differences in the comprehension of expository text. He also has developed a computer-based assessment tool for working memory span. Joe Magliano has done extensive research on assessing reading strategies during narrative comprehension and the use of verbal protocols to reveal these strategies. Magliano and Millis have recently conducted research regarding the use of LSA to predict reading skill based on think-aloud protocols. Their research provides a basis for the developing the procedures used by the proposed tutor to assess reading skill and the online assessment of the effectiveness of the training. Finally, Katja Wiemer-Hastings has worked for 3 years with Art Graesser to develop and test his Autotutor. As such, she has expertise regarding automated tutors, developing procedures for providing appropriate feedback to users with LSA, and LSA database management. Collectively, these six researchers possess all of the qualifications required to successfully complete this project.

CONCLUSION

The purpose of the IERI initiative is to "improve preK-12 student learning and achievement in reading, mathematics, and science by supporting rigorous, interdisciplinary research on large-scale implementations of promising educational practices and technologies in complex and varied learning environments." To this end, an important goal of the IERI Initiative is to facilitate the accumulation of reliable and valid data. We have strove to meet the high standards necessary to achieve that goal. Specifically, our proposed experiments are methodologically sound and rigorously address our empirical questions. Moreover, the investigators are well-versed in scientific methodology and have more than sufficient experience within these research domains. In addition, two well-known educational researchers (Drs. Art Graesser, and Susan Goldman) have agreed to serve as consultants. Their primary role will be to preview our experimental methodologies in detail. This process will ensure unbiased, expert judgements of every aspect of this research. The second benchmark of the IERI initiative is that the project be of sufficient scale. A large-scale project is necessary in our case to answer our questions, and to do so in an experimentally rigorous fashion. Too often, research such as this is conducted by cutting corners due to insufficient funds. This grant will allow us to fully and rigorously tackle these complicated questions. Moreover, the inclusion of four school systems within three states will ensure a careful examination of how differing conditions within a variety of educational systems along with diverse set of students interact with the benefits of reading interventions. The third benchmark of this initiative is to integrate technology. The integration of technology with this project is integral to our goals. The recent advancements made with LSA have allowed us to more clearly conceive of ways to automate reading strategy training. By providing an automated system, we are also optimizing the conditions for our lessons learned to help needy students. Finally, we have met the benchmark of conducting our research as an interdisciplinary team. We have brought together psychologists, educational researchers, and computer scientists from two universities to collectively tackle these issues. Together, we believe that we will make a significant contribution to the needs of our educational system.

TIME LINE

RESEARCH AREA	**Year 1**	**Year 2**	**Year 3**	**Year 4**	**Year 5**
Measures					
Course Domain Knowledge measure validation	**X**	**X**			
Text Domain Knowledge measure validation	**X**				
Text Comprehension measure validation	**X**				
Question 1: Metacognitive Knowledge					
Experiment 1	**X**	**X**			
Questions 2 and 3: Reading Intervention Effectiveness					
Preparation and preliminary tests of interventions	**X**				
Experiment 2 (Pilot study)		**X**			
Experiment 3			**X**		
Experiment 4				**X**	
Question 4: Automated Reading Strategy Instruction					
Computer programming	**X**	**X**	**X**		
Protocol analyses to validate LSA evaluations	**X**	**X**	**X**		
Usability tests for automated tutor		**X**	**X**		
Experiment 5		**X**	**X**		
Experiment 6				**X**	**X**
Experiment 7				**X**	**X**