Promoting Active Reading Strategies to Improve Students' Understanding of Science

Activities (Year 1: October 2000 to June 2001)

READING STRATEGY TRAINING

The 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). Our goals this year concerning the empirical examination of reading strategy training were the following:

1. Conduct a Pilot study (i.e., a modification of Experiment 2 in the proposal) to compare the three programs for reading strategy training: Insert, Previewing, and Self-explanation Reading Training (SERT).
2. Conduct a Baseline study (i.e., a modification of Experiment 1 in the proposal) at the four primary school sites to establish baseline measures of skill and knowledge for their science courses.

To accomplish the first goal (i.e., Pilot study), we achieved the following objectives:

1. We established contacts with two local schools, which were not among our four primary sites. These two schools were chosen because they were ethnographically heterogeneous with relatively large minority populations (i.e., 51% and 37% minority; 18% and 13% free or reduced price lunch). The teachers at the two chosen schools agreed to have their Earth Science classes participate and permission was obtained by the Chesapeake, VA, school superintendent.
2. We developed a prior knowledge measure for the earth science course. This involved an in-depth *knowledge analysis* of the first three chapters of three earth science textbooks. The most common concepts needed to understand the texts were identified and test questions were included to target those concepts. The majority of the questions were obtained from state learning tests, though some were experimenter-created (all multiple choice). We then piloted the test with college students to identify unclear and exceptionally difficult questions.
3. We created the Insert and Previewing training programs. This involved writing scripts for each program and piloting the programs with college students.
4. We redesigned the SERT program to be more suitable for high-school students.
5. We created the following measures: a metacognitive reading strategy checklist, a demographics questionnaire, mid-semester and final questionnaires to assess strategy use, and comprehension questions for five science texts.
6. We conducted the Pilot study with 16 classrooms (n=341) in the spring semester. The schools operated on a Block system meaning that the courses began in January and ended

in June with 90-minute classes. The design of this study and lessons-learned are described in the following *Findings* section.

7. We are currently obtaining the participating students' science grades and SOL scores from the schools.
8. We are currently scoring and entering the data.

To accomplish the second goal (i.e., Baseline study), we achieved the following objectives:

1. We established contacts with the four primary schools sites. This first involved contacting the schools' administration and obtaining permission to conduct our study (support from the superintendents was established during the proposal process). We then met with the administrators and teachers at each school (Americus, GA; Prestonburg, KY; Norfolk, VA; Williamsburg, VA) to discuss the Baseline study and to introduce the studies to be conducted the following year.
2. We developed a teacher practice survey to assess teachers' teaching practices with regard to reading and student strategy development. This survey was administered to the teachers participating in both the Pilot and Baseline studies (36 surveys were administered and 33 have been returned thus far).
3. We developed a prior knowledge test to assess knowledge of Air Masses (the topic of a science text read by the students) and general science knowledge (a subset of the original prior knowledge questions used in the pilot study).
4. We revised the demographics questionnaire.
5. We conducted the Baseline study at the four school sites with a total of 88 classrooms including 1,890 students in the following courses: Physical Science (n=14 classrooms), Earth Science (n=21); Chemistry (n=11); Biology (n=36); Physical Science/Biology (n=6). The design of this study and lessons-learned are described in the following *Findings* section.
6. We are currently obtaining the participating students' science grades and SOL scores from the schools.
7. We are currently scoring and entering the data. Our goal is to have entered all of the data for the Pilot and Baseline study by July 10 such that we can use this data to prepare for our experimental studies to be conducted in the fall. (This is, by the way, a lofty goal.)

AUTOMATED TRAINING PROGRAM DEVELOPMENT

An important goal of this research program is to develop computer software that administers reading strategy training to students. Such a training program involves a minimally interactive 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 then evaluates individuals' responses using recent technology provided by Latent Semantic Analysis (Landauer, Foltz, & Laham, 1998). We envision that tutoring program will consist of four major components: Assessment, Instruction, Modeling, and Practice. The first component assesses user needs for training. The particular methods of assessment will be guided by the results of our experiments and thus work on this component has not yet commenced. The second multimedia component illustrates and explains the key reading strategies, as identified by the earlier described experiments. The third interactive component involves the students' viewing strategy use, and then identifying which strategies were used. The fourth component allows the user to practice the strategies while reading texts. The computer evaluates the responses and provides fundamental feedback and hints when appropriate. The activities concerning the latter three components are described below.

<u>Development of the Instructional Component</u>

Our goals for the first year of the project in regard to the development of the instructional component were:

1. Develop a version of the trainer, which functions without LSA technology, to deliver SERT training. This development allows us to determine requirements, evaluate potential implementations, and, in the second year, determine user acceptability. This trainer is expected to have minimal interactive components, augmented when the LSA module is available.
2. Develop a version of the trainer, which functions without LSA technology, to deliver Previewing training. This trainer will be functionally useful in the future in developing a more extensive reading strategy trainer. In addition, since the Previewing methodology is conceptually simpler than SERT, the Previewing trainer, in various phases of development, serves as a prototype for solving technical problems encountered during the development of the SERT trainer.

To accomplish these goals we achieved the following objectives:

1. We developed requirements for the automated trainer. We determined that it should follow the general outline of the human-based training but need not follow that pattern slavishly: The automated trainer can provide an enhanced experience since it is not constrained by a group setting. Furthermore, the trainer should capture the trainee's interaction with the software to a database for the evaluation of student progress and analysis of statistical data. In addition, we determined that trainer instruction should be via the internet for the following reasons: ease of delivery to many recipients, ease of upgrading and augmenting services, and ease of data collection for analytical purposes. Finally, given that the student audience for the trainer may contain many individuals with low motivation, the trainer should be attractive, engaging and supportive.
2. We investigated several technologies as candidates for the delivery of the automated trainer. To do so, we built several prototypes to assess usability and acceptability. These prototypes included streaming video, streaming pictures accompanied by streaming audio, and animated characters with a text-to-speech engine. After building several small prototypes, we selected the animated characters as the technology to employ. We rejected streaming video for several reasons: it requires a bandwidth that may not be available in many school systems in the near term; it is hard to employ in a responsive interactive fashion; and it presents a significant development problem in that each new idea requires a small video production. The streaming picture solution does not have the bandwidth problem to the same degree, but the accompanying streaming audio frequently lacks smoothness over constrained bandwidths. A more significant problem is that user acceptance was perceived to be low. The succession of single frames seems to create a

distance between presenter and audience and the lack of moving lips is a significant negative.

In accordance with these findings, we chose to use animated characters because they do not have the bandwidth limitation, they convey a sense of a real (although cartoonish) person, and they provide a facile experimental designing solution. While the characters require large files to be installed on the users' machines, these files need be downloaded only once and can be pre-delivered via CDROM. Then, the entire process can be controlled programmatically from ordinary web pages, which are virtually all text and therefore quickly downloaded even from dial-up lines. The characters have facial expressions, their lips move, and they boast a repertoire of gestures that enliven the presentation. From a development point of view the technology is extremely flexible. To revise a script requires only textual editing, not re-shooting or re-recording. Characters are easily programmed to respond to user interaction. The text-to-speech engine seems to be acceptable to users in connection with these agents due to their somewhat cartoonish quality whereas it is less acceptable in conjunction with streaming pictures of a real person.

3. We have partitioned the project into five major parts so that development can proceed in parallel. The first three parts derive from the basic design of the trainers. The trainers proceed in three stages. First we introduce the strategy techniques. This stage corresponds to the introductory lecture in the human-based training. However, to make this more acceptable to the trainees and to motivate user response, we employ several interacting animated characters. The second stage is modeling. During this stage, the trainees see animated characters using the reading strategies; then, they engage in a question and answer period about the kind of strategies used. The third stage is practice. The trainees employ the strategies and the trainer responds with encouragement and direction. The final two parts being developed are the creation of our own characters and the integration of data collection into the trainer.

4. We are developing our own animated characters. We have designed four characters, a preliminary list of their gestures, and are acquiring appropriate props for them. Once completed, we will plug them into our existing scripts with little modification necessary.

5. We developed the trainers’ introductions to SERT and Previewing. The problems dealt with were both artistic and technical. We mastered the technologies needed to coordinate the characters interacting with each other and with the presentation of material via the web browser. Since we are using several characters, the introductions take the form of plays rather than lectures. We experimented with two types of plays: a reporter interviewing an expert, and a teacher providing instruction to one or two students. We chose the latter. We distributed the presentation information to several characters, giving each a personality, and assessed the presentation for effectiveness and acceptability. At this point we are polishing the introductory presentations. Although we broke the presentations into navigable segments, we would like to give the user the option of "rewinding" the presentation with a finer grain of control. This remains a problem to be solved. In addition, when our own characters have been developed we will revise the segments to accommodate them.

6. We are developing the modeling stage for the SERT trainer. This stage is patterned after the same stage in the human-based training and requires a good deal of interaction with the user. The user is shown an animated character representing a student giving self-explanations of sentences from a science text. The instructor character then asks the user to identify the types of self-explanation strategies used, where they were used within the self-explanation, and to which part of the text they apply. We developed a data structure that supports a text and its associated explanations and also developed several styles to use in asking these questions. We are in the process of adding a question manager, which takes account of the student's history in deciding among the styles of question to use. This part of the project is independent of the introduction part. We anticipate having a usable version by the end of the summer.
7. We are beginning development of the third part of the SERT trainer: providing practice to the user. In this phase, the trainee is given a science text to self-explain. The trainee types self-explanations into the trainer, which makes an appropriate response. When the LSA module is perfected, we will be able to use it to judge which types of strategy the trainee employed in the self-explanation and therefore the trainer will be able to respond appropriately. The trainer can then identify the strategy to the trainee or encourage the trainee to use additional strategies. In the meantime, we are substituting less sophisticated means of judging trainees' responses. These may use some combination of a less fine-grained evaluation of the trainee's response and constraining the trainee's response to something less open-ended than free-form text entry. We now possess a database of self-explanations to develop a test-bed for evaluating such algorithms. We anticipate having a usable version by the end of the first year of the project.
8. The last independent part of development is that of the database interface. The database must acquire all of the navigation, the questions asked, and the answers provided by the users. In addition, we intend to incorporate some indication of the reasoning used by the program at certain points (e.g., deciding to use a simpler method of asking questions because a student is not performing well). We separated the database development from the rest of the project and will integrate it once the other project modules are near finality. We established communication between the program and an Oracle database and can store the history of a trainee's navigation throughout the trainer pages. We are developing a database design to handle the more difficult problem of storing the questions and reasoning of the trainer and the trainee's responses. We decided that this database should store low-level information, which can be transformed at some later time into data suitable for analysis. By doing this, we can incorporate the database interaction into the library of procedures and functions used by the introduction, modeling, and practice modules developers.

<u>Development of the Latent Semantic Analysis Component</u>

In the "practice" component of the tutor, the student will read expository texts and type explanations to each sentence as instructed during the SERT training. Our overall goal is to build and use LSA to recognize students' reading strategies (i.e., paraphrasing, using logic and common sense, predicting, bridging, and/or elaborating).

In the basic architecture, LSA will be used to compare the student's input to "benchmarks" associated with each text sentence. The benchmarks are a collection of words that represent different types of thought that readers may use for a particular sentence. The contents of the benchmarks correspond to different types of reading strategies. LSA values across the benchmarks are expected to correspond to specific reading strategies. Building this architecture involves three important goals:

1. Build a suitable LSA text database.
2. Identify words that best reveal strategies via LSA.
3. Identify the patterns of LSA values across the benchmarks that best predict reading strategies.

The following objectives have been met with regard to the first goal of building the LSA database:

1. We constructed a text database on science, including texts from Biology, Physical Science, and Earth Sciences. The construction of the corpus is 90% complete.
2. We completed a preliminary evaluation of the text corpus via a keyword check to estimate the topic coverage. Approximately 80% of keywords were directly matched (i.e., without mediation through synonyms) by the text corpus.
3. We set up a systematic study to compare different methods of using LSA to identify reading strategies (paraphrase, elaboration, bridging, explanation, use of logic and commonsense). This study involves a sample of about 500 test sentences, and uses large preexisting LSA databases. (This study also addresses goal 3.)

To identify words that best reveal participants' explanation strategies via LSA, we completed the following study:

1. Two hundred and twelve college freshmen read two of four expository texts and typed in responses after each sentence. Some of these students were given SERT training beforehand, whereas others were trained to either paraphrase, elaborate and use common sense, predict, or bridge. This study produced approximately 4,800 responses. The students also took a comprehension test of the material that they had read. We are in the process of coding and analyzing this data.
2. We are building, coding and organizing the benchmarks associated with each sentence. We are coding one-half the collected benchmarks on strategy (paraphrase, prediction, etc.), the source of the words (the current sentence, a prior sentence, or world knowledge), and also the extent to which the response extends, elaborates, and explains the current sentence. The other half of the benchmarks will be used to test whether LSA, in conjunction with the first set of benchmarks, successfully identifies the strategies.

To accomplish our third goal of identifying patterns of LSA values across the benchmarks that best predict reading strategy, we have accomplished the following objectives:

1. We are in the process of identifying benchmarks and exploring ways that we can best predict reading strategies based on the pattern and the magnitude using LSA. Therefore, we are coding and organizing the corpus of verbal input ("the benchmarks") in a manner that will be flexible, suiting a number of research goals. One important constraint is that collecting large amounts of verbal protocols as benchmarks may be unfeasible when later adding new texts to the trainer. It is unclear whether the benchmarks will be

necessary for the tutor to recognize different reading strategies. Therefore, we are exploring various methods with which the tutor might recognize reading strategies.

2. The different approaches reflect the possibility that LSA may not be able to distinguish between all reading strategies (bridging vs. common sense vs. paraphrase, etc.). This seems likely because even human coders sometimes find the differences to be quite subtle. Indeed, the trainer might only differentiate relevant and irrelevant input. Although this outcome alone would not be ideal, it is essential for the trainer to have this capability.
3. We are adopting a two-pronged approach for coding, organizing, and using the benchmarks. The first – and more optimistic approach - will attempt to have LSA identify each of the SERT strategies. One-half the collected benchmarks will be coded on strategy (paraphrase, prediction, etc.), the source of the words (the current sentence, a prior sentence, or world knowledge), and the extent to which the response extends, elaborates, and explains the current sentence. The other half will serve as test cases. These data will enable us to examine whether different patterns of the source of the ideas correspond to different reading strategies. For example, paraphrases should be correlated with words from the current sentence, whereas using logic and common sense should be correlated with world knowledge and prior sentences.
4. The second approach is to build benchmarks primarily or solely from the words taken from the text. If successful, adding new texts to the trainer could be done with minimal work (in regard to building the benchmarks). For each sentence, we will identify: a) content words (nouns, verbs, adjectives, adverbs) which are new to the text so far, and b) content words that are old – they had been mentioned earlier in the text. These will serve as benchmarks, along with a group of words that imply reasoning (e.g., “so,” “because”, etc.).

Findings (Year 1: October 2000 to June 2001)

READING STRATEGY TRAINING

The following section describes the methodology for the Pilot and Baseline studies conducted this year, and the *lessons learned* during the process of completing these studies. The results from these studies will be available in August 2001.

Pilot Study

Method

The pilot study was conducted as a preliminary examination of the differences and effectiveness of the three training programs (see Experiment 2 in the proposal). The study included 341 students from two high schools with 9 Earth Science classrooms from Site A, and 7 Earth Science classrooms from Site B. There were 186 males and 155 females, of whom 53% were Caucasian and 40% were African-American.

There was a primary experiment, and two sub-experiments. The primary experiment was designed to compare the three training programs (Insert, Previewing, SERT) to a control condition. Three teachers from each site taught two regular (non-honors) Earth Science classes. Hence, one of each teacher's classes was randomly assigned to an experimental condition (Insert, Previewing, SERT), the other was randomly assigned to participate as a Control condition such that there were a total of 6 experimental and 6 control classrooms.

At Site A, the three teachers also each taught an Earth Science Honors class, hence using the same condition assignment as in the primary experiment, each Honors class participated in one of the three experimental conditions (this secondary experiment lacks a control condition). At Site B, one of the three teachers taught a third Earth Science course which was assigned to the SERT condition to allow a sub-comparison with one teacher between Insert, SERT, and the Control condition.

For the primary experiment and the two sub-experiments, the methodology was the same. The experiment included 6-7 class sessions including a pretest; training (1-2 sessions); a 1-week delay post-test; 2 mid-semester reminders; and a final post-test at the end of the year. The pretest session occurred the second week of classes. We administered the Gates-MacGinitie (Form K), a Science Prior-Knowledge test, a multiple-choice reading strategy test, a reading strategies checklist, and a demographics questionnaire. The following week, participating students were provided with reading strategy training. The SERT and Previewing conditions required two class periods (about 2 hours), whereas the Insert and the Control conditions required 1 class period. The control condition consisted of reading the same texts as those read in the training conditions and discussing topic questions. At the end of the training session, the students read a science text (a chapter about Glaciers) and answered comprehension questions. Students in the experimental conditions were instructed to use the strategy they had learned during training. One week following training (post-test), the students read a science text (a chapter about Air Masses), answered comprehension questions, and completed a reading strategies checklist to indicate reading strategies used when reading the text. Students in the experimental conditions were instructed to use the strategy they had learned during training. After a 4-week and 10-week interval, students in the experimental conditions were provided with a brief reminder for the reading strategy they had learned, and all participating students were given a questionnaire to assess reading frequency and strategy use. During a final post-test session (at the end of the semester), students completed the Gates-MacGinitie (Form L), the multiple-choice reading strategies test, the reading strategies checklist, and a demographics questionnaire. They also read two science texts and answered questions about the texts.

Results

We are currently scoring and entering the data from this study. We expect to be able to report results from this study by August.

Baseline Study

Method

The purpose of the Baseline study was two-fold (see Experiment 1 in the proposal). First, it is a correlational study to examine the relationships between reading skill, reading strategy knowledge, science knowledge, science text comprehension, and science course performance. Second, it provides baseline measures for the four sites to be included in our study next year (i.e., Experiment 3 in the proposal). We conducted the Baseline study at the four school sites (Norfolk, VA; Williamsburg, VA; Prestonburg, KY; and Americus, GA) with a total of 88 classrooms including 1,890 students, of whom approximately 57% were Caucasian and 35% were African-American. The students were in the following courses: Physical Science (n=14 classrooms), Earth Science (n=21); Chemistry (n=11); Biology (n=36); Physical Science/Biology (n=6). We administered the Gates-MacGinitie (Form L), a brief Science Prior-Knowledge test, a multiple-choice reading strategy test, a reading strategies checklist, and a demographics questionnaire. They also read a text about Air Masses and answered comprehension questions concerning the text. The session required 90 minutes. For the two schools with 90-minute class sessions, testing required one session. For the two schools with 50-minute class sessions, testing was administered across two sessions.

Results

We are currently scoring and entering the data from this study. We expect to be able to report results from this study by August.

Lessons Learned

Although our quantitative data is not yet ready to present, our pilot study was also geared toward refining our testing and training procedures. In that respect, it was time well spent – our lessons learned during the past year will serve us well in the upcoming year. The following is a list of our most important 'lessons learned'.

1. Previewing was much more difficult for the students to learn than we expected. We need to simplify our training protocol and include more scaffolding.
2. If possible, we need to reduce the number of tests administered during the pretest (a grueling session).
3. Training was disruptive to the teachers' schedule because we covered different topics during training than what the teacher was covering at that time. We need to use texts during training such that we are covering the topics that they are covering in class. This approach sacrifices the experimental rigor of using the same text for all training conditions, but will allow a more natural application of the reading strategies for the students.
4. Refresher training needs to be provided using classroom texts and maintaining teachers' curriculum plans.
5. Communication with teachers (and administrators) is problematic. We have now adopted a three-pronged approach including phone calls, faxes, and letters.
6. Collecting the students' grades at the end of the year has been difficult. Although the schools agreed to provide them to us at the beginning of the study, a couple of schools (in both the Pilot and Baseline studies) refused when it came time to do so (they have since complied, but with reluctance). In future studies, we will obtain a written agreement with each school and request, in advance, a volunteer teacher to collect the data. This individual will be reimbursed for time spent in this endeavor.

AUTOMATED TUTOR DEVELOPMENT

Development of the Instructional Component

The development of the instructional component involved a series of decision-making tasks regarding the automated trainer's requirements and specifications. The activities, resulting conclusions, and implementations were presented in the *Activities* section.

Development of the Latent Semantic Analysis Component

The development of the LSA practice component has involved a combination of computer programming tasks and empirical studies. The programming activities, conclusions, and implementations were presented in the *Activities* section. The results from the empirical investigations will be available in October 2001.