iSTART: Interactive strategy training for active reading and thinking

DANIELLE S. McNAMARA
University of Memphis, Memphis, Tennessee

and

IRWIN B. LEVINSTEIN and CHUTIMA BOONTHUM
Old Dominion University, Norfolk, Virginia

Interactive Strategy Training for Active Reading and Thinking (iSTART) is a Web-based application that provides young adolescent to college-age students with high-level reading strategy training to improve comprehension of science texts. iSTART is modeled after an effective, human-delivered intervention called self-explanation reading training (SERT), which trains readers to use active reading strategies to self-explain difficult texts more effectively. To make the training more widely available, the Web-based trainer has been developed. Transforming the training from a human-delivered application to a computer-based one has resulted in a highly interactive trainer that adapts its methods to the performance of the students. The iSTART trainer introduces the strategies in a simulated classroom setting with interaction between three animated characters—an instructor character and two student characters—and the human trainee. Thereafter, the trainee identifies the strategies in the explanations of a student character who is guided by an instructor character. Finally, the trainee practices self-explanation under the guidance of an instructor character. We describe this system and discuss how appropriate feedback is generated.

Unfortunately, many students have difficulty understanding what they read, particularly in challenging textbooks such as those used in science courses (Bowen, 1999; Snow, 2002). Some of these difficulties stem from students' lack of knowledge about and failure to use metacognitive comprehension strategies while reading their course material (Cottrell & McNamara, 2002; Cox, 1997, Garner, 1990; O'Reilly & McNamara, 2002). Many of these readers have sufficient word and syntactic decoding skills to read the words and sentences but lack the ability to understand the texts at the deep, meaningful level required to do well in their courses (see, e.g., Gough, Hoover, & Pe-terson, 1996). Our Web-based system, called the *Interactive Strategy Training for Active Reading and Thinking* (iSTART), addresses this problem by teaching young adolescent through college-age students to use a variety of reading strategies.

iSTART is modeled after a human-delivered reading strategy intervention called *self-explanation reading training*, or SERT (McNamara, in press; McNamara & Scott, 1999). SERT was inspired by research showing the benefits of self-explanation, which is the process of explaining texts or problems to oneself (see, e.g., Chi, Bassok, Lewis, Reimann, & Glaser, 1989; Chi, de Leeuw, Chiu, & LaVancher, 1994). Self-explanation enables readers to engage actively with a text and thus process information at a deeper level. However, some readers do not naturally self-explain texts and self-explain poorly when prompted. Therefore, SERT was designed to improve students' ability to self-explain difficult text by combining self-explanation training with metacognitive reading strategy training.

Metacognitive reading strategies are increasingly recognized as critical to successful, skilled reading. Skilled readers are more likely to engage in comprehension monitoring and active reading strategies such as previewing, predicting, making inferences, drawing from background knowledge, and summarizing (Brown, 1982; Long, Oppy, & Seely, 1994; Magliano & Millis, 2003; Oakhill, 1984; Oakhill & Yuill, 1996). Moreover, instruction to use reading strategies improves readers' ability to comprehend text (Baker, 1996; Baumann, Seifert-Kessell, & Jones, 1992; Bereiter & Bird, 1985; Bielaczyc, Pirolli, & Brown, 1995;

Davey, 1983; Dewitz, Carr, & Patberg, 1987; Hansen & Pearson, 1983; Palinscar & Brown, 1984; Yuill & Oakhill, 1988).

SERT teaches readers to self-explain text and to use metacognitive reading strategies that improve self-explanation and, consequently, comprehension and learning. The intervention coaches trainees in five reading strategies. The underlying strategy is *comprehension monitoring*, or being aware of how well one understands while reading. Research has shown that learners who score high on measures of metacognition and comprehension monitoring generally outperform learners who score low on these measures (Pressley & Ghatala, 1990). Comprehension monitoring enables the reader to recognize a failure of understanding, and it is this recognition that triggers the use of additional active reading strategies.

SERT recommends that the first of these active strategies to employ is restating the sentence in the reader's own words, or *paraphrasing*. Although paraphrasing by itself does not constitute an explanation, it helps to ensure that the reader understands the sentence's grammar and vocabulary. Paraphrasing may be enough to trigger comprehension; it may also alert the reader that some words are unfamiliar and that their meanings need to be deduced from context, discovered from later material in the text, or looked up. Paraphrasing can also help memory for the surface text structure by transforming the text into more familiar ideas. Information that is more familiar is more memorable.

SERT encourages students to go beyond this basic sentence-focused processing by invoking knowledge-building strategies that link the content of the sentences to other material. They are taught to use a strategy called *bridging* to tie the current sentence to the material previously covered in the text. Research has shown that better comprehenders are more likely to make bridging inferences (e.g., Oakhill, 1984). Such inferences allow the reader to form a more cohesive global representation of the text content (see, e.g., Kintsch, 1998).

Students may also use *prediction* to anticipate the content of the rest of the text, either by guessing what is coming next or by reminding themselves to watch out for some particular item that will aid comprehension. Though infrequently used, prediction has been found to be helpful in promoting text comprehension (see, e.g., Hansen & Pearson, 1983), particularly when predicted outcomes are highly probable (see, e.g., Magliano, Dijkstra, & Zwaan, 1996).

Finally, students may associate the current sentence with their own related prior knowledge gained from sources outside the text, using a strategy called *elaboration*. Research has established that both domain knowledge and elaboration are associated with improved learning and comprehension (e.g., Bransford & Johnson, 1972; Pressley et al., 1992; Spilich, Vesonder, Chiesi, & Voss, 1979). Elaboration essentially ensures that the information in the text is linked to information that the reader already knows. These connections to prior knowledge result in a more coherent

and stable representation of the text content (see, e.g., Kintsch, 1998). Readers are also encouraged to elaborate the text using domain-general knowledge, logic, or common sense, particularly when they do not have sufficient knowledge about the topic of the text.

The underlying assumptions of SERT are that instruction to use more effective reading strategies will improve readers' ability to self-explain text and, in turn, that the externalization of the reading strategies within the self-explanations renders the strategies more tangible to the reader. McNamara and Scott (1999; McNamara, in press) found that when readers were asked to self-explain a text, those provided with training used more effective strategies and understood the text better than those who did not receive training. Moreover, reading strategy training was most effective for students who had less prior knowledge about the domain of the text. Readers with more prior knowledge about a topic generally understand more from a text (see, e.g., McNamara, 2001; McNamara, E. Kintsch, Songer, & W. Kintsch, 1996; McNamara & W. Kintsch, 1996). In contrast, low-knowledge readers, particularly those who do not use active reading strategies to compensate for a lack of domain knowledge, often understand text poorly (Cottrell & McNamara, 2002; O'Reilly & McNamara, 2002). However, SERT provides students with strategies that they can use to compensate for their lack of domain knowledge. Thus, although low-knowledge students are less able to use domain-specific knowledge to make sense of the text and to fill in conceptual gaps (see, e.g., McNamara et al., 1996), they are able to use domain-general knowledge strategically to enhance their understanding. Indeed, an analysis of students' self-explanation protocols has verified that low-knowledge readers who were provided with SERT learned to use their common sense and logic to understand the difficult text (McNamara, in press).

The three phases of SERT training (introduction, demonstration, and practice) can be administered to a small group of students in about 2 h. The training begins with a brief introduction that includes definitions and examples of self-explanation and the reading strategies. In the demonstration section that follows, the trainees are given a text and shown a video of a student reading and self-explaining it aloud. The video is interrupted at several points, and the trainees are asked to identify the strategies that the student in the video was using. The student in the video models the desired self-explanation behavior and the discussion of strategies clarifies points that had been presented more abstractly during the lecture. The demonstration gives the trainees concrete instances of the behaviors introduced in the lecture. Assisted by instructors, the students then work with partners to practice the strategies, taking turns in reading and explaining the text.

Assessments of SERT (as delivered by human trainers) have established that SERT is remarkably effective, resulting in substantial increases in both comprehension and exam scores (e.g., McNamara, in press; McNamara,

Best, & Castellano, 2004; McNamara & Scott, 1999). However, the cost of delivering this reading strategy intervention using human trainers is relatively high. Thus, it would be challenging, if not impossible, to provide SERT to a wide audience. iSTART was developed so that SERT could be made available to a large number of students at minimal cost. Automating SERT offers additional advantages, such as self-paced learning and individualized training. In this article, we discuss the transformation of a classroom-oriented training program into an electronic, one-on-one trainer. In the sections that follow, we describe the iSTART system, including its design considerations, its pedagogy, the algorithms used to evaluate trainees' explanations, and the accuracy of the system's evaluations during training in comparison with evaluations made by human judges.

iSTART Trainer: Design Considerations

iSTART was created to provide SERT in a more scalable and cost-effective manner than is possible using human trainers. Some of the constraints on the project were the same as those that limit human-based training. Most importantly, the training had to be designed as a relatively brief intervention that could be readily adaptable to a variety of science disciplines. The reason for the requirements of both brevity and adaptability is that high-school science teachers have a great deal of material to cover during the year. It is difficult to persuade them to adopt a training program that takes class periods away from lessons focused on course material. Teachers are generally more willing to accept a program that involves material on the subject they are teaching. In addition, we needed to be able to make frequent revisions to the trainer and to collect detailed data on the interactions of the trainees with the software. For this reason, iSTART was designed to be accessed via an Internet browser from a central server that could record all student interactions in a database.

In designing iSTART, our goal was to preserve the basic character of the human-based training, build on its strengths, and adapt it to the Web environment. The basic character of the original training lies in its three phases: (1) an *introductory* explanation that includes definitions and examples, (2) a *demonstration* of the techniques in action, and (3) an opportunity to *practice* the techniques under guidance. A major strength of this training is the variety of ways in which the trainees interact with the SERT strategies: They receive them as a passive audience for definitions and examples, see them in use by someone like themselves, analyze those uses in a group discussion, observe and prompt another trainee who is using the techniques, and practice the techniques themselves. We wanted to maintain the phases and variety of the human-based training when we developed iSTART for the web environment. In that environment, the computerized trainer would deal with one trainee at a time, rather than with a classroom. The trainers in the classroom gained authority from the regular classroom teacher

and from their position in front of the class; iSTART, in contrast, would have no such advantage. In addition, trainees with computer experience have expectations of the Web environment that are far different from their expectations of the classroom. Finally, human trainers and fellow trainees are intelligent enough to respond appropriately to attempts at self-explanation made during the practice sessions, an ability which takes a good deal of programming in a computer application.

Although we wanted to maintain the qualities of the human-based training, we also wanted to take advantage of the web environment, which afforded many opportunities for improving the human-led classroom training. Because it deals with trainees one at a time, the training can be self-paced and adaptive to the trainee. More important, the training can engage the trainee in more interactive dialog and require the trainee to be a more active learner. In developing the trainer, we faced two major problems: (1) how to maintain the trainees' interest in and involvement with the training process while still being effective (the pedagogical problem) and (2) how to determine the quality of the trainees' explanations so that the trainer could respond appropriately (the evaluation problem). The pedagogical problem required the replacement of passive aspects of the human-based training with interactivity appropriate to a computer-based setting. The evaluation problem was complicated by the fact that, rather than simply determining whether an answer was correct or incorrect, we had to recognize a trainee's typed response as constituting one strategy or another or as exhibiting a quality.

iSTART Trainer: Animated Pedagogy

The human-based training motivates SERT, introduces its basic concepts, and trains students in its techniques using interactive discussion and practice. The challenge to replace this human interactivity was met by the use of animated agents. Our use of animated agents to provide training was inspired by a system called AutoTutor, developed by Art Graesser and colleagues (e.g., Graesser, Wiemer-Hastings, Wiemer-Hastings, Kreuz, & the Tutoring Research Group, 1999). AutoTutor employs a *talking head*, which interacts with a student-user who is learning course content such as Newtonian physics (see, e.g., Graesser et al., 2003). The constraints on iSTART, however, differ from those on AutoTutor in at least two significant ways. First, AutoTutor is concerned with the accuracy of the content of the trainees' responses rather than with their use of strategies. This issue is discussed in a following section. Second, the human-led training begins with about 30 min of lecture regarding self-explanation and reading strategies. In designing iSTART, we assumed that 30 min of a talking head lecture would fail to engage the user sufficiently to allow successful learning of the concepts. The human-based SERT training is delivered in a classroom and uses group-oriented lectures and demonstrations. Ordinary classroom discipline and social pressure generally suffice to keep trainees' attention during these presenta-

tions. In a computer-based environment, more than a talking head is required to maintain attention for long periods of time.

Introduction. Our approach to this problem during the introduction phase was to simultaneously take advantage of vicarious learning (see, e.g., Shebilske, Jordan, Goettl, & Paulus, 1998) and create an engaging, interactive learning environment via dialogue and question asking (see, e.g., Craig, Gholson, Ventura, Graesser, & the Tutoring Research Group, 2000; Graesser, Hu, & McNamara, in press). Observing a dialogue between several people who are interacting with each other has been shown to benefit learning (see, e.g., McKendree, Stenning, Mayes, Lee, & Cox, 1998). Hence, to increase interactivity and more fully engage the student, we made use of a trio of animated agents to deliver the introduction module. In addition, because it was odd to have three heads interact on a monitor, we used full-body figures with exaggerated heads so that mouth movements and facial expressions would be more visible.

The characters we designed have a variety of gestures and speak using a text-to-speech synthesizer. As they speak, their words appear in a balloon above their heads. The only drawback in this system is that the characters and text-to-speech engine must be installed on the trainees' computers. However, most current personal computers include all of the software needed apart from the iSTART pedagogical agents. The required components can be downloaded and installed quickly with only a basic knowledge of Microsoft Windows. After the initial installation, no additional installation is necessary because the characters' behavior is controlled by scripts embedded in the web pages delivered to the browser from the server. Any mod-

ifications to iSTART are made to pages that reside on the server.

In the iSTART introduction module, we replace the human-delivered lecture with a classroom discussion among an instructor character, two student characters, and the trainee (see Figure 1). This format provides the trainee with a greater variety of interactions with the concepts and strategies than does the human-delivered training. These interactions take the following forms: (1) The instructor character presents material as in a lecture with both definitions and examples; (2) the instructor character questions the student characters; (3) the student characters banter among themselves; (4) the student characters ask the instructor character for examples or clarifications; and (5) trainees answer questions that assess and remediate their understanding of the concepts.

The introduction is broken down into a number of sections, and the trainee has to click on a button to move from one to the next. In addition, the trainee is given brief quizzes at several points during the introduction and advised to continue to the next section or to review the last, depending on the results. By all of these means, the trainee is induced to be an active participant in the presentation. There is little opportunity for the trainee's attention to wander and, should the trainee desire, he or she can easily return to an earlier part of the introduction for review.

Demonstration. The theme of interactivity is carried into the demonstration module of iSTART (see Figure 2). In the human-based training, the training group is shown a video of a student self-explaining a text. The video is interrupted at several points for a discussion concerning which strategies were used in the explanation. The same
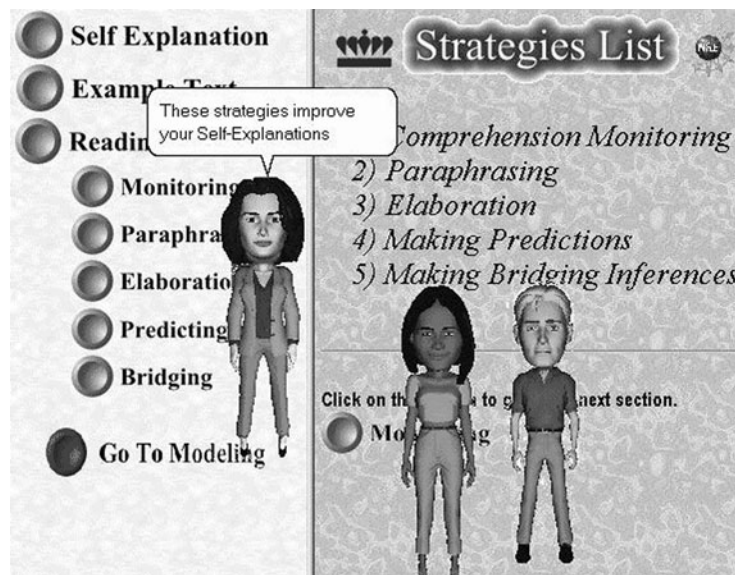

Figure 1. Screen shot from the SERT introduction module.

basic approach is taken in iSTART: The text is read and explained sentence by sentence by a student character (Genie) under the supervision of a teacher character (Merlin). After Genie produces a satisfactory explanation, the teacher asks the trainee to decide which strategies were used in Genie's explanation. At times, Merlin will ask follow-up questions or give instructions that focus the trainee's attention on the explanation and the target sentence.

The demonstration module dynamically adapts to the success of the trainee by varying the style of questioning that Merlin uses. The trainee's success is measured by the number of correct answers provided in the last few interactions. If the success level is low, more supportive styles of questioning are used. If it is high, less follow-up is required. The range of follow-up styles, from most to least supportive, includes: (1) Tell the trainee that a certain strategy was used, remind the trainee of the definition of that strategy, and ask the trainee to identify where a certain strategy is used by clicking on the text of the explanation; (2) as in (1), but omit the reminder of the definition; (3) ask the trainee to decide which strategies on a list of choices were used; usually ask the trainee to find two different strategies; ask follow-up questions in detail; and (4) as in (3), but ask fewer follow-up questions.

Trainees usually begin the demonstration module by being questioned in multiple-choice style, as in (3) above. This is a moderately difficult task, since several strategies may be used in one explanation. In this mode, the trainees have to discriminate among strategies while they discriminate among sections of the explanation. If they correctly identify a strategy used in the text, they may be given one or two follow-up questions or instructions,

such as "Click on the part of the explanation where that technique was used" or, in the case of the bridging strategy, "Click on the part of the text to which this explanation linked." Merlin may also ask the trainee to identify another strategy or give a prompt that a certain strategy was used and then ask the trainee to locate it. If they are not successful, the trainer switches to the less demanding modes, in which the trainees are told that a particular strategy is used and then asked to locate its position in the explanation. Trainees who are successful with the multiple-choice questioning with respect to longer explanations are presumed to have a good understanding of the techniques. Once they reach this level, it is not necessary to verify their understanding by persisting in the follow-up, so fewer and fewer follow-ups are used as long as the trainees continue successfully.

Evaluation of Trainee Responses: Practice

In the demonstration module, the trainees have analyzed the self-explanations of a surrogate trainee (i.e., Genie) who has used the full range of reading strategies. Thus, the demonstration section provides the trainee with a mental model of self-explanation that they can use when they are asked to explain a text in the practice module. In human-based training, trainees are paired up and asked to coach each other as they practice explanations of a selected passage of a science text. In iSTART, the teacher character (i.e., Merlin) interacts with the trainee. The interaction follows the pattern that the trainee has observed in the demonstration section. During the practice module, the trainee self-explains sentences from texts while attempting to use the reading strategies learned in the introduction and demonstration sections.
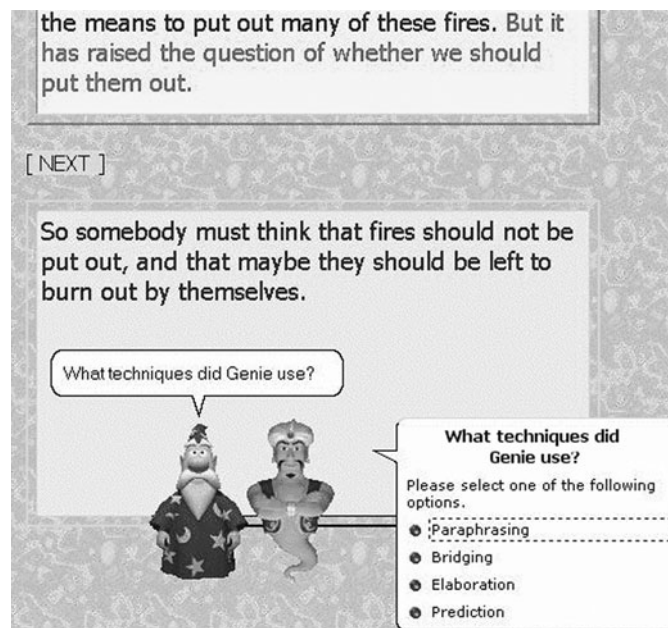

Figure 2. Screen shot from the SERT demonstration module.

The computational challenge faced during this phase is to provide appropriate feedback to the trainee on the basis of the quality of the self-explanations. We have approached this "evaluation" challenge in three ways. First, the explanation is *screened* to determine whether it is sufficiently long, and sufficiently different from and relevant to the sentence and/or topic. If any of these three criteria is not met, then the trainee is asked to modify the explanation by adding more information or more details. Second, the trainee's explanation is *evaluated* in terms of its quality in order to guide the feedback provided by the teacher agent. Third, the trainee is directly *asked* what strategies were used. In the following three subsections, we describe the algorithms that we have used to guide these three approaches. We then present data describing the accuracy or appropriateness of these evaluations in comparison with human evaluations.

Initial Screening

The primary goal of the iSTART system is to move the reader away from simply restating the sentence and toward adding relevant information to explain the content. Hence, the system first screens a response to examine its length relative to the sentence, how closely it matches the sentence in words and length, and whether it contains words that are relevant to the sentence or the text topic. These algorithms are based on several factors: (1) the content words in the target sentence, including nouns, verbs, adjectives, and adverbs; (2) associated words for each content word that were identified manually on the basis of collected protocols and dictionaries; and (3) a length priority manually assigned to each sentence, which partially determines how long an explanation must be in order to be counted as long enough and which was based on the length of the sentence and analyses of prior protocols collected for those sentences; and (4) a strategy priority indicative of the importance of the sentence, which was based on text and protocol analyses.

The strategy priority determines whether to do a follow-up. The first three factors are used to determine the quality of the student's explanation. Specifically, the words in the self-explanation are matched against the content words in the target sentence and against the association lists for each of the content words. This is accomplished in two ways: (1) by exact matching of the first 75% of the characters, and (2) by matching based on a *soundex* (see, e.g., Knuth, 1998) transformation of each word. An exact soundex match is required for short words (i.e., those with fewer than six characters) due to a high number of false alarms when soundex is used. For longer words, a match on the first four soundex symbols suffices. Soundex is used to compensate for misspellings by dropping vowels and mapping similar characters (e.g., *t* and *d*, *s* and *z*) to the same soundex symbol. The matching process yields the number of exact and soundex matches to important and associated words. The algorithms also make use of the word counts for the target sentence and for the entire explanation.

Explanations are categorized as too short if the length of the self-explanation is less than the product of the length of the sentence and the length priority. For example, if the length of the sentence is 10 words and the length priority is 1, then the required length of the self-explanation would be 10 words. If the length of the sentence is 30 words and the length priority is 0.5, then a passing self-explanation would require 15 words. If a self-explanation is categorized as too short, Merlin responds to the trainee, "Hmm, your explanation is kind of short. Could you expand your answer? I like details."

An explanation is considered irrelevant if it does not contain at least two words related to the sentence, matching either content words from the target sentence or words from the association lists corresponding to the content words. Merlin's response to an irrelevant explanation is, "Please try to add more information that is related to the sentence."

An explanation is categorized as too similar if (1) it contains 60% of content words in the target sentence, (2) its length is 80% to 120% of the length of the target sentence (in terms of number of words), and (3) it contains fewer association words than content words. Merlin's response in this case is, "Try adding some more information that explains what the sentence means."

Overall Evaluation

If the trainee's explanation passes the screening tests, it is evaluated in terms of its quality. The three levels of quality that guide feedback to the trainee are based on two factors: the number of matched words in comparison with the number of content words in the target sentence, and the length of the explanation in comparison with the length of the target sentence. Self-explanation length guidelines for each sentence were established by examining a set of explanations previously collected from participants who were not using the iSTART system. Cutoffs were set to reflect the quality of those explanations. An explanation meets the criteria for Level 3 if it is more than 220% of the target sentence length or has more relevant (i.e., matched) words than the number of content words in the target sentence. Merlin's response in these cases is "Excellent!" "Superb!" or "That's really great!" Level 2 criteria are met if the explanation is more than 145% of the target sentence's length, to which Merlin responds "Good job," "That's fine," or "That's pretty good." Otherwise, the explanation is categorized as Level 1, in which case Merlin responds with "O.K.," "Try saying more next time," or "Let's try the next one."

Asking the Trainee

The third method that we use to evaluate the trainee's responses entails a more direct strategy: We simply ask the trainees what strategy they have used. With this direct approach, we achieve two goals. First, we are encouraging the trainee to engage in metacognitive self-evaluation. This forces the trainee to think about the self-explanation more objectively and to think more

about the different types of strategies that might have been used. Second, by using this approach we will be able to assess retrospectively the ability of both the trainees and our system to categorize self-explanations (i.e., by comparing the trainees' and the system's assessments with experimenter ratings).

Throughout the interaction, the module tracks the variety of strategies employed. Selected sentences have been identified as particularly appropriate for certain strategies. When they come to one of these sentences, trainees may be encouraged to use a particular strategy if it has not already been used during the practice.

The aim of the practice module is to give the trainee an opportunity to employ the strategies while self-explaining a difficult text. Having the trainee identify the strategies used in his or her own self-explanations is simply a means to that end, but too much interaction slows the trainee's progress from one sentence to another. Therefore, the module gradually adopts a less intrusive posture than that of the demonstration section, during which every sentence was explained and every explanation analyzed. At first, iSTART asks students to identify the strategies used in their self-explanations and continues with a follow-up. If practice proceeds successfully, the teacher character becomes less directive, perhaps just prompting the trainee to go on to the next sentence or occasionally asking the trainee to expand an explanation that seems too short or irrelevant. In this way, the trainee can get on with the main task of this section, which is, of course, practice.

Evaluation: A Comparison With Human Judges

A primary interest of ours is the success of our algorithms in evaluating student responses. These algorithms first screen responses that are too short, too similar to the sentence being explained, or apparently unrelated to that sentence. If no problems are identified, they then evaluate the quality of the sentence on a scale of 1 to 3. To judge the appropriateness of these evaluations, we compared the evaluations made by the system with human-based judgments for self-explanation protocols elicited from 42 middle-school children using iSTART. Because these protocols were collected during training, they had already been screened by the system. Thus, the children's final self-explanations (i.e., after Merlin's prompts and feedback) for each sentence of a 13-sentence text about thunderstorms were evaluated by human judges. These human judgments were compared with the iSTART evaluation scores, which ranged from 1 to 3 (see, also, Best, Ozuru, & McNamara, in press).

First, the explanations were categorized by the human judges as containing (1) only a paraphrase, (2) an elaboration that is not directly relevant to the comprehension of either the current sentence or the overall text, (3) a relevant elaboration that contributes to the comprehension of only the target sentence, and (4) a relevant elaboration that contributes to a more global level of comprehension of the text than just the current sentence (e.g., actively building the large picture depicted by the overall text).

The reliability of the coding frame was analyzed using Cohen's kappa and simple agreement analysis. In all cases, reliability between the coders for each dimension was 85% or above. Disagreements were resolved via a discussion between the coders.

To assess the accuracy of the iSTART evaluation system, we examined the iSTART score as a function of human classifications. The cross-tabulations of these data are presented in Table 1. We also computed the average iSTART score as a function of category.

First, we used analysis of variance (ANOVA) to examine whether the iSTART score successfully distinguished between those explanations containing only paraphrases ($n = 316$) and those containing some form of elaboration ($n = 224$). Indeed, there was a significant difference in iSTART scores between paraphrases and elaborations [$M = 1.89$ and $M = 2.22$, respectively; $F(1,538) = 51.7$]. This result indicates that the responses to users tended to distinguish appropriately between explanations in terms of quality. Notably, it is also evident in Table 1 that only 28 paraphrases (9%) were misclassified as Level 3 explanations.

With reference only to elaborations, our second goal was to examine whether iSTART successfully distinguished among the three categories of irrelevant ($n = 45$, $M = 2.18$), sentence relevant ($n = 154$, $M = 2.21$), and globally relevant ($n = 25$, $M = 2.36$). Although the trends were in the right direction, these differences were not reliable ($F < 2$). It is important to note, however, that global elaborations were never misclassified by iSTART as Level 1 explanations (see Table 1). Thus, those instances never received inappropriate feedback during training. For those that were categorized as sentence relevant, only eight explanations, or 5%, were misclassified as Level 1. We consider these results to be very encouraging.

Finally, we were interested in whether the iSTART score discriminated between elaborations that were based solely on prior text (i.e., bridging inferences) and those based on knowledge (either general or domain specific). One possibility was that iSTART would misclassify elaborations that are based on knowledge (as opposed to text) as poor or irrelevant explanations. This would occur if the association list, which is the primary source of identifying relevancy aside from the sentence itself, did not adequately anticipate words that a trainee might use in an explanation. If this were the case, then self-explanations based on information outside the text (and our associa-

Table 1
iSTART Evaluation Scores as a Function of the Category of Elaborations as Judged by Human Coders

| iSTART Score | Human-Judged Categories of Elaborations | | | | |
	Paraphrase	Irrelevant	Sentence	Global	Total
1	60	2	8	0	70
2	228	33	105	16	382
3	28	10	41	9	88
Total	316	45	154	25	540

tion list) would not be recognized as relevant to the sentence. This would be an unfortunate outcome, given that our primary goal is to get the readers to use their knowledge to go beyond the text and, thus, explain it.

To examine iSTART's success in this regard, the students' elaborations were coded as having their source in the text, in general knowledge, or in scientific knowledge. Note that this was an iterative categorization; those explanations containing all three sources would be classified as scientific knowledge. On the basis of these classifications, an ANOVA indicated that the difference between iSTART scores as a function of source was marginal [$F(1,221) = 2.6, p = .07$]. This result reflected marginally higher scores for text-based elaborations ($M = 2.37$) than for either elaborations from general knowledge ($M = 2.18$) or those from scientific knowledge ($M = 2.17$).

The cross-tabulations of these human classifications with the iSTART scores are presented in Table 2. It is notable that there are substantially more elaborations based on general knowledge than on either the text or scientific knowledge. Of course, those elaborations that were based on general knowledge most likely also included information from the text (since this was an iterative classification). Hence, this result indicates that the majority of the participants were relying on the prior text and general knowledge to explain the sentences. This result may be reflective of the knowledge levels of the young population targeted in this study. That is, their prior knowledge of science was relatively sparse.

Nonetheless, it is notable that a relatively small percentage of the knowledge-based elaborations were assigned to Level 3 by iSTART (i.e., 24% of general knowledge, 21% of scientific knowledge). In contrast, 38% of the text-based elaborations were scored as Level 3. This indicates that the system is slightly less likely to highly rate a knowledge-based explanation in comparison with a text-based elaboration. We discuss our current attempts to remediate this situation more fully in the Discussion section.

Completion Times

Average total completion times for the three iSTART modules are presented in Table 3 for two experiments. In the first experiment, iSTART was completed by 72 college students who practiced with a 20-sentence passage about thunderstorms. The second experiment included 42 middle-school students. Modifications of iSTART for

the experiment with the middle-school students included a slight shortening of the introduction and the addition of a second practice text. Specifically, the middle-school students were given a 13-sentence version of the thunderstorms passage ($M = 26.66$ min, $SD = 8.00$) and a 12-sentence passage about the origin of coal ($M = 24.45$ min, $SD = 7.07$). Initial instructions and instructions for the separate modules (presented by Merlin) required approximately 10 min. Hence, iSTART training requires approximately 1.5–2 h.

DISCUSSION

Our goal is to address what we view as a critical need in our educational system: to provide a large number of students with reading strategy training based on empirically supported and theoretically grounded reading strategy research. We believe that iSTART is a sound foundation for meeting those needs. Our approach was to develop an automated training system utilizing pedagogical agents in an interactive, engaging interface. One challenge that we have faced is how to provide appropriate feedback to our trainees during self-explanation practice. Our results thus far are both encouraging and indicative of needs to improve our system.

The results are encouraging because iSTART rarely gives inappropriate feedback to the trainee. Those explanations categorized by human judges as relevant elaborations were misclassified as Level 1 (poor but sufficient) explanations only eight times. In addition, the probability of appropriately classifying a paraphrase as Level 1 was four times the probability of doing so with elaborations. Hence, when a trainee is sticking too closely to the sentence, the system is relatively good at giving appropriate feedback. This result is heartening given the relative simplicity of our word-based algorithms used to evaluate the self-explanations.

It seems, however, that iSTART was not able to identify irrelevant explanations (i.e., 43 explanations categorized by humans as irrelevant were given scores of 2 or 3 by iSTART). However, there are two considerations that should be taken into account. First, these protocols included only those explanations that had already been prescreened for irrelevancy (because the trainees were using iSTART at the time). Hence, we do not have a clear estimate of hits and false alarms in this regard. Second, responding in an overly enthusiastic manner only 8% of

Table 2
iSTART Evaluation Scores as a Function of the Source of Elaborations as Judged by Human Coders

| iSTART Score | Human-Judged Source of Elaborations | | | |
	Prior Text	General Knowledge	Scientific Knowledge	Total
1	1	8	1	10
2	31	105	18	154
3	20	35	5	60
Total	52	148	24	224

Table 3
Average Completion Times (in Minutes) for iSTART Across Two Experiments

		Introduction		Demonstration		Practice	
Experiment	N	M	SD	M	SD	M	SD
College-age subjects	72	37.80	4.4	12.04	1.99	29.98	5.96
Middle-school subjects	42	29.81	10.10	11.61	2.15	51.11	7.59

the time seems almost human. That is, we would rather walk on the encouraging side than be overly critical, risking frustration on the part of the trainee.

In this regard, observations of and reports by the students indicate that they enjoy using the system. Virtually no frustration has been observed while participants engage with iSTART. Moreover, frustration is not evident in participants' self-explanation protocols. In addition, our questionnaire data indicate that participants like using the system. For example, the children in this study gave an average rating of 4 (out of 5) to the helpfulness of the practice section and to how much they had learned from the system. In sum, participants tend to be happy with their experiences using iSTART.

We would like, however, to do a better job identifying explanations based on the trainee's prior knowledge (both general and domain specific). We would also like to have a more flexible system—one that can be used with any text on short notice. The use of association lists restricts both of these goals. Therefore, we need a system that is relatively independent of the text itself yet adequately representative of the external knowledge a trainee may bring to bear regarding a text. Indeed, latent semantic analysis (LSA) appears to have this capability. LSA is a theory and method for extracting and representing a universe of discourse using statistical computations applied to a large corpus of text (Landauer & Dumais, 1997; Landauer, Foltz, & Laham, 1998). The corpus embodies a set of mutual constraints that largely determine the semantic similarity of words and sets of words. These constraints can be resolved using linear algebraic methods, particularly singular value decomposition. For our purposes, LSA can be used to assess the semantic similarity between sets of words with reference to the corpus, but without being explicitly bound to the exact words used in the text being explained (Landauer & Dumais, 1997). Among a large range of applications, LSA has been used to grade essays (Foltz, Gilliam, & Kendall, 2000), predict text comprehension (Shapiro & McNamara, 2000), and provide feedback in AutoTutor (see, e.g., Graesser et al., 2000). Indeed, Millis et al. (2004) present some initial research on various methods of using LSA to classify self-explanations. On the basis of this work, we believe that a combination of our word-based algorithms and LSA techniques will improve iSTART's ability to provide appropriate feedback to the user.

Our approach in this initial system has been to provide general feedback to the trainees. Now that we have established that the feedback is generally on target, our future versions will incorporate more specific advice to users. For example, when the system identifies explanations as being too close to the original sentence, it will advise thinking about how the sentence relates to previous sentences or to what the trainee already knows. Indeed, current research in our laboratory is focused on the relative advantages of specific versus general feedback, as well as on the effects of its appropriateness.

Does iSTART effectively improve students' ability to comprehend texts? Although not the focus of this article, this is certainly an important question. Our laboratory results from three experiments conducted thus far are highly positive in this regard (McNamara & the CSEP Lab, 2004; O'Reilly, Sinclair, & McNamara, 2004). First, a study of 300 college students was conducted to compare iSTART, SERT (the live version), and a control condition (in which the students read the same texts but were not given strategy instruction). The results confirmed that iSTART is as effective as SERT in terms of improving the quality of students' self-explanation in comparison with that of the self-explanation of the control participants. In addition, the students read a science text and answered comprehension questions 1 week after training. The results showed reliable advantages for both iSTART and SERT in comparison with the control condition (O'Reilly et al., 2004).

A second study was conducted with 42 middle-school students (McNamara & the CSEP Lab, 2004). In this study, half of the students were provided with training and half were not before they were asked to read and self-explain a text about heart disease. The comprehension data indicate that the locus of gains from iSTART training depended on both the students' prior knowledge of reading strategies and the level of comprehension assessed (see, e.g., Kintsch, 1998; McNamara et al., 1996). Specifically, children with less prior knowledge of reading strategies who received iSTART performed better on text-based questions than did their counterparts in the control condition. Thus, less strategic children gained primarily in terms of understanding the text at the text-base level of comprehension. In contrast, those with more prior knowledge of reading strategies showed greater gains from iSTART training in terms of developing a deeper understanding of the text as assessed by bridging inference questions.

This pattern of results was also found in a third study of 44 college students (McNamara & the CSEP Lab, 2004). These students were asked to read and self-explain two texts before receiving iSTART training and to do the same with two texts after training. As we found with the middle-school students, the results of this study indicated that better readers gained in terms of deeper levels of comprehension. That is, they performed better on bridging inference questions after training than before. In contrast, less skilled readers gained in terms of their surface-

level understanding of the text, showing significant gains on text-based questions.

In sum, our results thus far indicate that students will make progress in their area of proximal development (see, e.g., Vygotsky, 1978). It appears that readers first need to learn to form an adequate representation of the text-based information—essentially, the information presented in each individual sentence. Then, readers can learn how to understand the text at a deeper level by processing the relationships between the ideas conveyed across sentences. iSTART allows this progression of improvement by providing training at both levels of processing. However, these results indicate that less skilled readers would benefit from more extensive training than that provided in the iSTART version described here. In future studies, we will examine whether more extensive training can bridge the gap between more and less skilled comprehenders.

We are currently conducting a year-long study on whether high-school students gain from iSTART in comparison with a control group (which received training to create Web pages), in terms of both immediate gains, in comprehension and course grades after training, and longer-term gains at the end of the school year. In this study, we are testing iSTART within the classroom environment so that it can eventually be integrated with high-school classroom curricula. Along these lines, we would like to note that iSTART is in its infancy in terms of the scope we envision. Currently, iSTART provides SERT only within a relatively restricted domain (science). Future versions are anticipated in which training is provided for a broader range of strategies (at both lower and higher levels) and domains (e.g., history, literature, social studies) in a way that more flexibly accommodates the abilities of the trainees. Expanding iSTART in these ways will increase the likelihood that it can be used in a variety of classrooms. On one hand, the increased flexibility will make it useful to more students. Those with lower verbal abilities may need training at lower levels (e.g., word identification and paraphrasing), whereas students with higher verbal abilities may benefit from going beyond SERT—for example, by learning to adapt strategy use to the context. On the other hand, increasing the range of text domains should increase the likelihood that iSTART will be used in the classroom. Teachers will be able to assign texts or topics for self-explanation training that are being covered in class, reducing the time taken away from course material and thereby improving students' understanding of the course topic.[1]

iSTART was inspired and is guided by constructivist theories of comprehension and learning that emphasize the active use of knowledge. In line with these theories, McNamara (1997; McNamara, de Vega, & O'Reilly, in press; McNamara & McDaniel, 2004) has described a knowledge-based account of comprehension skill. The knowledge-based account of reading comprehension skill asserts that skilled comprehenders more actively and efficiently use prior knowledge to comprehend text than unskilled comprehenders do (see also Bereiter & Bird, 1985; Ericsson & Kintsch, 1995; MacDonald & Christiansen, 2002; McNamara & Scott, 1999, 2001; Paris & Jacobs, 1984; Snow, Burns, & Griffin, 1998). This knowledge use may arise either because they have more prior knowledge about the text topic (see, e.g., McNamara & McDaniel, 2004) or because they know and use more metacognitive reading strategies (see, e.g., McNamara & Scott, 1999, 2001).

The underlying assumption of the knowledge-based account (McNamara, 1997; McNamara et al., in press; McNamara & McDaniel, 2004) is that greater activation and use of knowledge, rather than more stable mechanisms such as working memory capacity or suppression mechanisms (see McNamara et al., in press) drives reading comprehension skill. According to the knowledge-based account, training less skilled comprehenders to use strategies that promote knowledge use and integration should improve their comprehension. Although this outcome seems intuitive, it is not an a priori prediction on the basis of other models of reading comprehension skill (see, e.g., Gernsbacher, 1990; Just & Carpenter, 1992; Rosen & Engle, 1998; cf. Kintsch, 1998). Hence, the success of iSTART in improving comprehension ability supports the knowledge-based account over other accounts of reading comprehension skill.

iSTART surpasses our expectations in many respects. Most importantly, the students we have observed using iSTART are completely engaged in the learning process. This contrasts with our human-delivered version of SERT, in which students had many more chances for wandering attention. Since SERT trainees are members of a group, they are under less pressure to respond actively than the iSTART trainees. Essentially, iSTART is more intensive than its SERT counterpart. Students' engagement results partially from the use of animated characters. However, we believe that it also emerges from the consistent incorporation of cognitive principles of learning in the design and implementation of iSTART.

REFERENCES

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

Baumann, J. F., Seifert-Kessell, N., & Jones, L. A. (1992). Effect of think-aloud instruction on elementary students' comprehension monitoring abilities. *Journal of Reading Behavior*, **24**, 143-172.

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

Best, R., Ozuru, Y., & McNamara, D. S. (in press). Self-explaining science texts: Strategies, knowledge and reading skill. *Proceedings of the Sixth International Conference of the Learning Sciences*.

Bielaczyc, K., Pirolli, P. L., & Brown, A. L. (1995). Training in self-explanation and regulation strategies: Investigating the effects of knowledge acquisition activities on problem solving. *Cognition & Instruction*, **13**, 221-252.

Bowen, B. A. (1999). Four puzzles in adult literacy: Reflections on the national adult literacy survey. *Journal of Adolescent & Adult Literacy*, **42**, 314-323.

Bransford, J., & Johnson, M. K. (1972). Contextual prerequisites for

understanding some investigations of comprehension and recall. *Journal of Verbal Learning & Verbal Behavior*, **11**, 717-726.

BROWN, A. (1982). Learning how to learn from reading. In J. A. Langer & M. T. Smith-Burke (Eds.), *Reader meets author: Bridging the gap* (pp. 26-54). Newark, DE: International Reading Association.

CHI, M. T. H., BASSOK, M., LEWIS, M. W., REIMANN, P., & GLASER, R. (1989). Self-explanations: How students study and use examples in learning to solve problems. *Cognitive Science*, **13**, 145-182.

CHI, M. T. H., DE LEEUW, N., CHIU, M., & LAVANCHER, C. (1994). Eliciting self-explanations improves understanding. *Cognitive Science*, **18**, 439-477.

COTTRELL, K., & McNAMARA, D. S. (2002). Cognitive precursors to science comprehension. In W. D. Gray & C. D. Schunn, *Proceedings of the Twenty-Fourth Annual Meeting of the Cognitive Science Society* (pp. 244-249). Mahwah, NJ: Erlbaum.

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

CRAIG, S., GHOLSON, B., VENTURA, M., GRAESSER, A. C., & THE TUTORING RESEARCH GROUP (2000). Overhearing dialogues and monologues in virtual tutoring sessions: Effects on questioning and vicarious learning. *International Journal of Artificial Intelligence in Education*, **11**, 242-253.

DAVEY, B. (1983). Think aloud: Modeling the cognitive processes of reading comprehension. *Journal of Reading*, **27**, 44-47.

DEWITZ, P., CARR, E., & PATBERG, J. (1987). Effects of interference training on comprehension and comprehension monitoring. *Reading Research Quarterly*, **22**, 99-121.

ERICSSON, K. A., & KINTSCH, W. (1995). Long-term working memory. *Psychological Review*, **102**, 211-245.

FOLTZ, P. W., GILLIAM, S., & KENDALL, S. A. (2000). Supporting content-based feedback in online writing evaluations with LSA. *Interactive Learning Environments*, **8**, 111-129.

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

GERNSBACHER, M. A. (1990). *Language comprehension as structure building*. Hillsdale, NJ: Erlbaum.

GOUGH, P. B., HOOVER, W. A., & PETERSON, C. L. (1996). Some observations on a simple view of reading. In C. Cornoldi & J. Oakhill (Eds.), *Reading comprehension difficulties: Process and intervention* (pp. 1-13). Mahwah, NJ: Erlbaum.

GRAESSER, A. C., HU, X., & McNAMARA, D. S. (in press). Computerized learning environments that incorporate research in discourse psychology, cognitive science, and computational linguistics. In A. F. Healy (Ed.), *Experimental cognitive psychology and its applications: Festschrift in honor of Lyle Bourne, Walter Kintsch, and Thomas Landauer*. Washington, DC: American Psychological Association.

GRAESSER, A. C., JACKSON, G. T., MATHEWS, E. C., MITCHELL, H. H., OLNEY, A., VENTURA, M., CHIPMAN, P., FRANCESCHETTI, D., HU, X., LOUWERSE, M. M., PERSON, N. K., & THE TUTORING RESEARCH GROUP (2003). Why/AutoTutor: A test of learning gains from a physics tutor with natural language dialog. In R. Alterman & D. Hirsh (Eds.), *Proceedings of the Twenty-Fifth Annual Conference of the Cognitive Science Society* (pp. 1-5). Mahwah, NJ: Erlbaum.

GRAESSER, A. C., WIEMER-HASTINGS, K., WIEMER-HASTINGS, P., KREUZ, R., & THE TUTORING RESEARCH GROUP (1999). AutoTutor: A simulation of a human tutor. *Journal of Cognitive Systems Research*, **1**, 35-51.

GRAESSER, A. C., WIEMER-HASTINGS, P., WIEMER-HASTINGS, K., HARTER, D., PERSON, N., & THE TUTORING RESEARCH GROUP (2000). Using latent semantic analysis to evaluate the contributions of students in AutoTutor. *Interactive Learning Environments*, **8**, 129-147.

HANSEN, J., & PEARSON, P. (1983). An instructional study: Improving the inferential comprehension of good and poor fourth-grade readers. *Journal of Educational Psychology*, **75**, 821-829.

JUST, M. A., & CARPENTER, P. A. (1992). A capacity theory of comprehension: Individual differences in working memory. *Psychological Review*, **99**, 122-149.

KINTSCH, W. (1998). *Comprehension: A paradigm for cognition*. New York: Cambridge University Press.

KNUTH, D. (1998). *The art of computer programming* (2nd ed., Vol. 3). Reading, MA: Addison-Wesley.

LANDAUER, T. K., & DUMAIS, S. T. (1997). A solution to Plato's problem: The latent semantic analysis theory of acquisition, induction, and representation of knowledge. *Psychological Review*, **104**, 211-240.

LANDAUER, T. K., FOLTZ, P. W., & LAHAM, D. (1998). Introduction to latent semantic analysis. *Discourse Processes*, **25**, 259-284.

LONG, D. L., OPPY, B. J., & SEELY, M. R. (1994). Individual differences in the time course of inferential processing. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **20**, 1456-1470.

MACDONALD, M. C., & CHRISTIANSEN, M. H. (2002). Reassessing working memory: Comment on Just and Carpenter (1992) and Waters and Caplan (1996). *Psychological Review*, **109**, 35-54.

MAGLIANO, J. P., DIJKSTRA, K., & ZWAAN, R. (1996). Generating predictive inferences while viewing a movie. *Discourse Processes*, **22**, 199-224.

MAGLIANO, J. P., & MILLIS, K. K. (2003). Assessing reading skill with a think-aloud procedure and latent semantic analysis. *Cognition & Instruction*, **21**, 251-283.

McKENDREE, J., STENNING, K., MAYES, T., LEE, J., & COX, R. (1998). Why observing a dialogue may benefit learning. *Journal of Computer Assisted Learning*, **14**, 110-119.

McNAMARA, D. S. (1997). Comprehension skill: A knowledge-based account. In M. G. Shafto & P. Langley (Eds.), *Proceedings of the Nineteenth Annual Meeting of the Cognitive Science Society* (pp. 508-513). Mahwah, NJ: Erlbaum.

McNAMARA, D. S. (2001). Reading both high-coherence and low-coherence texts: Effects of text sequence and prior knowledge. *Canadian Journal of Experimental Psychology*, **55**, 51-62.

McNAMARA, D. S. (in press). SERT: Self-explanation reading training. *Discourse Processes*.

McNAMARA, D. S., BEST, R., & CASTELLANO, C. (2004). Learning from text: Facilitating and enhancing comprehension. Retrieved February 2004 from www.speechpathology.com.

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

McNAMARA, D. S., KINTSCH, E., SONGER, N., & KINTSCH, W. (1996). Are good texts always better? Interactions of text coherence, background knowledge, and levels of understanding in learning from text. *Cognition & Instruction*, **14**, 1-43.

McNAMARA, D. S., & KINTSCH, W. (1996). Learning from texts: Effects of prior knowledge and text coherence. *Discourse Processes*, **22**, 247-288.

McNAMARA, D. S., & McDANIEL, M. (2004). Suppressing irrelevant information: Knowledge activation or inhibition? *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **30**, 465-482.

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

McNAMARA, D. S., & SCOTT, J. L. (2001). Working memory capacity and strategy use. *Memory & Cognition*, **29**, 10-17.

McNAMARA, D. S., & THE CSEP LAB (2004). *Promoting active reading strategies to improve undergraduate students' understanding of science.* Annual project report submitted to the National Science Foundation IERI.

MILLIS, K., KIM, H.-J. J., TODARO, S., MAGLIANO, J. P., WIEMER-HASTINGS, K., & McNAMARA, D. S. (2004). Identifying reading strategies using latent semantic analysis: Comparing semantic benchmarks. *Behavior Research Methods, Instruments, & Computers*, **36**, 213-221.

OAKHILL, J. (1984). Inferential and memory skills in children's comprehension of stories. *British Journal of Educational Psychology*, **54**, 31-39.

OAKHILL, J., & YUILL, N. (1996). Higher order factors in comprehension disability: Processes and remediation. In C. Cornoldi & J. Oakhill (Eds.), *Reading comprehension difficulties: Processes and intervention* (pp. 69-92). Mahwah, NJ: Erlbaum.

O'REILLY, T., & MCNAMARA, D. S. (2002). What's a science student to do? In W. P. Gray & C. D. Shunn (Eds.), *Proceedings of the Twenty-Fourth Annual Meeting of the Cognitive Science Society* (pp. 726-731). Mahwah, NJ: Erlbaum.

O'REILLY, T., SINCLAIR, G. P., & MCNAMARA, D. S. (in press). Reading strategy training: Automated versus live. In *Proceedings of the 26th Annual Meeting of the Cognitive Science Society.*

PALINSCAR, A. S., & BROWN, A. L. (1984). Reciprocal teaching of comprehension-fostering and comprehension-monitoring activities. *Cognition & Instruction*, **2**, 117-175.

PARIS, S., & JACOBS, J. (1984). The benefits of informed instruction for children's reading awareness and comprehension skills. *Child Development*, **55**, 2083-2093.

PRESSLEY, M., & GHATALA, E. (1990). Self-regulated learning: Monitoring learning from text. *Educational Psychologist*, **25**, 19-33.

PRESSLEY, M., WOOD, E., WOLOSHYN, V., MARTIN, V., KING, A., & MENKE, D. (1992). Encouraging mindful use of prior knowledge: Attempting to construct explanatory answers facilitates learning. *Educational Psychology*, **27**, 91-109.

ROSEN, V., & ENGLE, R. (1998). Working memory capacity and suppression. *Journal of Memory & Language*, **39**, 418-436.

SHAPIRO, A. M., & MCNAMARA, D. S. (2000). The use of latent semantic analysis as a tool for the quantitative assessment of understanding and knowledge. *Journal of Educational Computing Research*, **22**, 1-36.

SHEBILSKE, W., JORDAN, J., GOETTL, B., & PAULUS, L. (1998). Observation versus hands-on practice of complex skills in dyadic, triadic, and tetradic training-teams. *Human Factors*, **40**, 525-540.

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

SNOW, C. E., BURNS, M., & GRIFFIN, P. (1998). *Preventing reading difficulties in young children*. Washington, DC: National Academy Press.

SPILICH, G., VESONDER, G., CHIESI, H., & VOSS, J. (1979). Text processing of domain-related information for individuals with high and low domain knowledge. *Journal of Verbal Learning & Verbal Behavior*, **18**, 275-290.

VYGOTSKY, L. S. (1978). *Mind in society: The development of higher psychological processes*. Cambridge, MA: MIT Press.

YUILL, N., & OAKHILL, J. (1988). Understanding of anaphoric relations in skilled and less skilled comprehenders. *British Journal of Psychology*, **79**, 173-186.

NOTE

1. iSTART is not yet available for public use. Teachers or their representatives should contact Danielle McNamara if they would like to be included in future experimental evaluations of iSTART. Researchers should also contact her concerning potential collaborations to evaluate experimentally iSTART. See http://csep.psyc.memphis.edu/istart/ for more information about iSTART.