

2007

iSTART 2: Improvements for Efficiency and Effectiveness

Irwin B. Levinstein
Old Dominion University

Chutima Boonthum

Srinivasa P. Pillarisetti

Courtney Bell

Danielle S. McNamara

Follow this and additional works at: https://digitalcommons.odu.edu/computerscience_fac_pubs



Part of the [Computer Sciences Commons](#), [Mathematics Commons](#), and the [Psychology Commons](#)

Repository Citation

Levinstein, Irwin B.; Boonthum, Chutima; Pillarisetti, Srinivasa P.; Bell, Courtney; and McNamara, Danielle S., "iSTART 2: Improvements for Efficiency and Effectiveness" (2007). *Computer Science Faculty Publications*. 88.
https://digitalcommons.odu.edu/computerscience_fac_pubs/88

Original Publication Citation

Levinstein, I. B., Boonthum, C., Pillarisetti, S. P., Bell, C., & McNamara, D. S. (2007). Istart 2: Improvements for efficiency and effectiveness. *Behavior Research Methods*, 39(2), 224-232. doi:10.3758/bf03193151

iSTART 2: Improvements for efficiency and effectiveness

IRWIN B. LEVINSTEIN

Old Dominion University, Norfolk, Virginia

CHUTIMA BOONTHUM

Hampton University, Hampton, Virginia

AND

SRINIVASA P. PILLARISETTI, COURTNEY BELL, AND DANIELLE S. MCNAMARA

University of Memphis, Memphis, Tennessee

iSTART (interactive strategy training for active reading and thinking) is a Web-based reading strategy trainer that develops students' ability to self-explain difficult text as a means to improving reading comprehension. Its curriculum consists of modules presented interactively by pedagogical agents: an introduction to the basics of using reading strategies in the context of self-explanation, a demonstration of self-explanation, and a practice module in which the trainee generates self-explanations with feedback on the quality of reading strategies contained in the self-explanations. We discuss the objectives that guided the development of the second version of iSTART toward the goals of increased efficiency for the experimenters and effectiveness in the training. The more pedagogically challenging high school audience is accommodated by (1) a new introduction that increases interactivity, (2) a new demonstration with more and better focused scaffolding, and (3) a new practice module that provides improved feedback and includes a less intense but more extended regimen. Version 2 also benefits experimenters, who can set up and evaluate experiments with less time and effort, because pre- and posttesting has been fully computerized and the process of preparing a text for the practice module has been reduced from more than 1 person-week to about an hour's time.

iSTART (interactive strategy training for active reading and thinking) is a Web-based reading strategy trainer that develops students' ability to self-explain difficult text as a way of improving reading comprehension. iSTART Version 1 was described in McNamara, Levinstein, and Boonthum (2004). The present article briefly reviews the background and design of that version and then reports the development of a revised version (Version 2). Our primary goals were to increase efficiency for the experimenters and to increase the effectiveness of the training. The modifications we applied were due to (1) deficiencies in pedagogy or efficiency discovered as we experimented with the first version in a high school setting and (2) the discovery of opportunities for improvements in pedagogy as we became more familiar with both the program and its audience.¹

iSTART is based on a human-delivered intervention called SERT (self-explanation reading training; McNamara, 2004b) that grew out of a think-aloud research tradition, which had established that explaining a text to oneself improves comprehension and that participants who produced better explanations also showed better comprehension (Chi, Bassok, Lewis, Reimann, & Glaser, 1989; Chi, de Leeuw, Chiu, & LaVancher, 1994).

SERT took these ideas a step further by training students in reading strategies to use when self-explaining a text (McNamara, 2004b; McNamara & Scott, 1999; O'Reilly, Best, & McNamara, 2004; O'Reilly, Sinclair, & McNamara, 2004b). The reading strategies include comprehension monitoring (noticing whether one understands the current sentence), paraphrasing (restating the sentence in one's own words), bridging (making connections between the current and prior sentences), elaboration (making connections between the current sentence and one's own knowledge or experience), and prediction (guessing what is going to happen in subsequent sentences). Experiments with SERT in first-year college science classes indicated that the training tended to compensate for lack of scientific background knowledge both in text-based comprehension (McNamara, 2004b) and in performance on course tests, including the final examination (McNamara, 2004a; McNamara & the CSEP Lab, 2005). The first version of iSTART was as successful as SERT with the same audience (Magliano et al., 2005; O'Reilly, Sinclair, & McNamara, 2004a; O'Reilly et al., 2004b).

iSTART is loosely modeled on the SERT curriculum and consists of three modules: (1) an interactive introduc-

I. B. Levinstein, ibl@cs.odu.edu

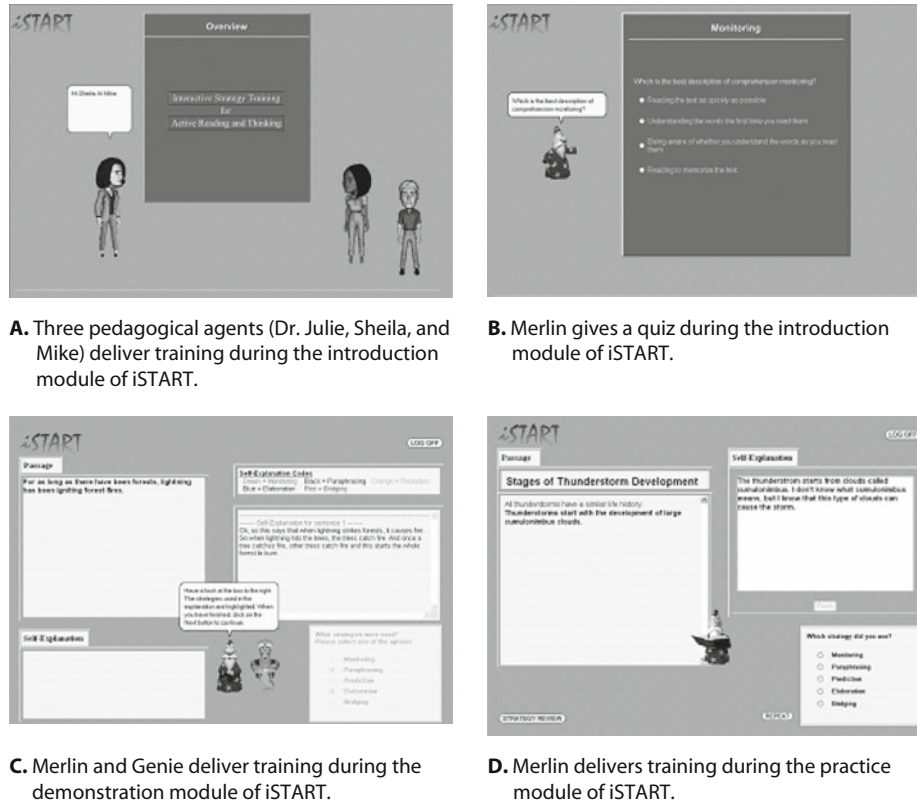


Figure 1. iSTART sample screen shots.

tion with three pedagogical agents who present and motivate the basics of self-explanation in a virtual classroom, a short quiz following each lesson; (2) a demonstration, in which the trainee analyzes one pedagogical agent's self-explanations of a science text while both the trainee and the first agent are coached by a second agent; and (3) a practice, in which the trainee generates self-explanations of other science texts under the guidance of the same pedagogical agent who coached the demonstration. Students also take a variety of pre- and posttests. Figure 1 shows some sample iSTART screen shots.

High School Experiment Experience

When tested in a high school environment, Version 1 of the trainer was measurably effective, as evidenced by a study including 39 eighth and ninth graders enrolled in a summer learning program (McNamara, O'Reilly, Best, & Ozuru, 2006). Students in the iSTART group performed significantly better than students in the control group, in terms of producing better quality self-explanations and higher comprehension scores. Results showed that iSTART students produced higher quality elaborations ($M = 1.11$, $SD = 0.79$) than control students ($M = 0.41$, $SD = 0.38$; Cohen's $d = 1.13$). In terms of comprehension, low-strategy-knowledge participants ($n = 20$) in the iSTART group ($M = .51$, $SD = .19$) performed significantly better than those in the control group ($M = .33$, $SD = .18$) on text-based questions [$t(18) = -2.26$, $p = .018$, Cohen's $d = 1.00$], whereas high-strategy-knowledge participants

in the iSTART condition ($M = .53$, $SD = .28$) performed better on bridging questions than did control participants ($M = .30$, $SD = .22$) [$t(18) = -1.90$, $p < .038$, Cohen's $d = 1.04$].

This study also proved beneficial in highlighting a number of the system's pedagogical problems and inefficiencies, as evidenced by researcher observations during the experiment, interviews with students and teachers, and answers to postexperiment questionnaires. Results from unpublished postexperiment questionnaire data from the study reported in McNamara et al. (2006) indicated an overall satisfaction with the first version of iSTART. Students were asked to rate various aspects of the system on a scale of 1 to 5 (low to high). Results are shown in Table 1. Students were also asked to specify what they did not understand about the system or what could make the system better. Their responses are categorized in Table 2.

The results suggest that when asked specifically about module helpfulness, module comprehensibility, and character likeability, students indicated an overall satisfaction, as shown by the average ratings for each question in Table 1. Students also indicated that they learned from the overall program. However, when asked in general about what they did not like or what could be improved, students reported some dissatisfaction with each module and each agent. Using the rating scale, students reported that they understood the introduction and that the introduction was helpful. However, responses to the open-ended questions indicated that students felt that more examples would help

Table 1
Mean Ratings (1 = Low to 5 = High) for
Student Responses to iSTART (Version 1) System Questions

Question	<i>M</i>	<i>SD</i>
Understanding of the introduction	3.750	0.809
The introduction's helpfulness in learning to use strategies	3.875	0.939
Understanding of the demonstration	3.575	0.958
Helpfulness of watching the demonstration	3.900	1.081
Helpfulness of the practice in learning strategies	4.025	1.074
How well strategy use was learned	4.025	0.800
How much was learned from the program overall	4.000	0.751
Frequency of future strategy use	3.575	1.059
Usefulness of the characters	3.225	1.349
Dr. Julie's likeability	2.875	1.343
Sheila's likeability	3.075	1.248
Mike's likeability	2.975	1.368
Merlin's likeability	3.875	1.202
Genie's likeability	3.600	1.317

them to better understand the strategies. Similarly, participants reported understanding the demonstration module, and said that watching the demonstration was helpful, but they also reported needing more help with the “forest fire” demonstration. Students also reported finding the practice helpful, but said that it was too long. Finally, although students reported that the agents were useful, they also said that the voices were annoying, and that “more real,” understandable characters were needed.

Because the high school experiments were much larger in scale than the college experiments, they magnified a number of inefficiencies in the system. Most importantly, experimenters found that the use of paper-based pre- and posttests necessitated extensive manual transcription, which introduced inaccuracies, including uncertainties as to which tests pertained to which students. Another inefficiency was revealed when the project's aims expanded to include nonscience disciplines. Use in a wider variety of classrooms would require the addition of many more practice texts, both because of the need to include more frequent practice sessions, as mentioned above, and because the teachers could more easily be recruited by including genres consonant with their specialties. Unfortunately, the process of preparing a text for the practice module in the initial version of iSTART required more than a person-week's time for textual analysis, dictionary construction, and preliminary experimentation.

In light of results from observations, questionnaires, and teacher and student interviews in conjunction with system inefficiencies, we made changes in creating the second version of iSTART. We added more examples to the introduction, to provide students with more opportunities to understand the strategies. We added an illustrative module (the “mini-demo”) to satisfy students' need for a preview of how the strategies would be used together in an explanation, to prepare them for the demonstration module, and to decrease the monotony of the introduction. We also provided more scaffolding in the demonstration in Version 2. Finally, we compensated for the practice's length by having the students self-explain only target sentences, rather than every sentence, and by extending practice over several weeks rather than having only a single session.

In what follows, we discuss the evolution of iSTART's second version in terms of the pedagogical changes made in the introduction, demonstration, and practice modules and the improved efficiencies that resulted from the modification of the explanation evaluation algorithm and the computerization of testing.

Changes to the Introduction Module

The first set of modifications to the original curriculum concerns the introduction. In the first version, the initial human-delivered lecture of the SERT intervention was replaced with a series of playlets involving three pedagogical agents (two students and a teacher) in a virtual classroom. The trainees observed their interaction and were given short quizzes on the content of the lessons after each playlet. We intended this to be a nonthreatening entry into the system, but it demanded too little of the students. Watching the scripted interaction was like watching a tele-

Table 2
Categorization of Student Responses to an Open-Ended
Postexperiment Question for iSTART (Version 1)
Concerning What Students Did Not Understand or
What Could Make the System Better

Introduction
Did not understand bridging, elaborating, and monitoring.
When they said to self-explain it, I did not get how to do it at first.
The monitoring and bridging strategies could have been explained more.
Self-explaining could have been explained better.
The strategy introductions could have been explained better.
I did not understand elaboration.
Give more examples for everything.
Give examples with the quizzes.
All of the reading strategies were sketchy.
Demonstration
The “forest fire” demonstration could have been more helpful.
Practice
I didn't like how long the practice was.
Agents
The voices were annoying.
I couldn't understand the characters.
Cut out the snoring.
“More real” characters.

vision program, and the quizzes simply notified the trainees whether they were right or wrong on each question.

We took several steps to increase active participation by the trainee. First, we added a mini-demo module, situated in the middle of the introduction section. The mini-module is designed to demonstrate and review how the first three strategies (comprehension monitoring, paraphrasing, and prediction) can be used together in a self-explanation. The mini-demo is modeled on the demonstration module that the trainee encounters after the introduction module. It shows a student character who is in the process of self-explaining a passage to a teacher character, and prepares the trainee for the actual demonstration module. This also adds an interaction in the middle of the introduction, as a change of pace.

Second, we increased the number of examples that illustrated the use of each of the strategies, and the teacher character encouraged the trainees to study them; they had to click to continue after each one. We also gave the trainees some control over the progress of the script, by adding a “pause” button. Finally, we reoriented the quizzes, so they changed from being mere assessments to being pedagogical opportunities. The students were given second and third chances to answer the questions, and were provided with instructional feedback when the initial answer was incorrect.

Changes to the Demonstration Module

The second set of modifications to the original iSTART curriculum concerns the demonstration module. In this module, the trainee examines the explanations produced by a pedagogical agent and tries to determine which strategies were used in the explanation, and where they were used. Each explanation uses several strategies, so the student is asked two questions per explanation. The questioning style for the first of the two is adaptive, based on a student-level variable that ranges from 1 to 9 and rises or falls as a function of the recent correct and incorrect answers of each student. All of the students begin at Level 5. At the lower levels, a student is told that a particular strategy is used and is asked to find it in the explanation, whereas at the upper levels, a student is asked to identify one of the strategies used (selecting from a list of the five strategies) and then is asked to find it. If the strategy is bridging, the student may be asked in a follow-up question to identify the sentence to which it was bridged. At the highest levels of the student-level variable, a student is asked fewer follow-up questions, whereas at the lowest levels, a student will be reminded of the definition of the strategy. For example: “Genie used paraphrasing in this explanation. You remember that paraphrasing means putting the sentence in your own words. Click on the part where Genie used paraphrasing.” Although the style of an explanation’s first question follows this adaptive format, the second question is always asked in multiple-choice fashion.

An analysis of the performance in Version 1 revealed alarming differences between college and high school students. From the data of one of the college experiments ($N = 56$), we calculated the number of students who were

ever rated at each of the levels of the student-level variable. Starting at Level 5, approximately half the students ever reached Level 4 and half ever reached Level 6. This balance continued at upper and lower levels, as can be seen in Figure 2A. More than a quarter of the college students reached Level 7. In contrast, among the high school students ($N = 176$; see Figure 2B), again starting at Level 5, fewer than a quarter ever reached Level 6, and only a tenth reached Level 7. Although only about 30% of the college students descended to Level 3, two thirds of the high school students did.

We suspect that the difference in success between the two groups is due to a wider variation in cognitive ability among the high school students. The eighth- and ninth-grade students in the summer learning program had been selected for their academic promise, and showed a performance profile on the demonstration module that was much more similar to Figure 2A than to Figure 2B. The task that we had set before the students requires a great deal of mental coordination, more than we had realized. To respond adequately, the student might need to consider the sentences in the explanation, the definitions of the strategies, the content of the current sentence, and the content of the prior sentences. Although the average college freshman was up to this task, the average high school student was not. An important consideration, too, is that the aim of the module is not assessment but training. Consequently, it was imperative to redesign the demonstration module to include much more scaffolding, with a question-management system that takes explicit account of the text used and the self-explanations offered by the pedagogical agent, rather than simply using the student-level variable, which did not factor in the complexity of either the selected sentence or its explanation. Especially at the lower levels, it was important to aid the student in achieving the insights needed to succeed at the higher levels.

Since the size and complexity of the task were the complicating factors, we addressed these in several ways. The requirement that the second question always include selecting a strategy from a list of five strategies was eliminated in favor of flexibility. Second, such lists could be reduced from five to as few as two choices, as shown in Figure 3. Third, the student could be directed to attend to portions of the self-explanations rather than having to review the entire explanation. Fourth, the portions focused on could be divided into sections showing where (but not which) different strategies were used, an option particularly important when a sentence of the explanation contains two or more strategies. These options are illustrated in Figure 4. We also made help more readily available, and when students identified the wrong location as the site of a particular strategy, they could be told which strategy was actually used.

One drawback of the modified version is that the new version of the demonstration imposes a high cost in time on the experimenters. The Version 1 module was completely automatic—the question style always varied in the same way, without regard to the difficulty or complexity of the sentence or explanation being analyzed. The new version provides a toolkit of techniques that have to be combined by the experimenter when designing a pair of question

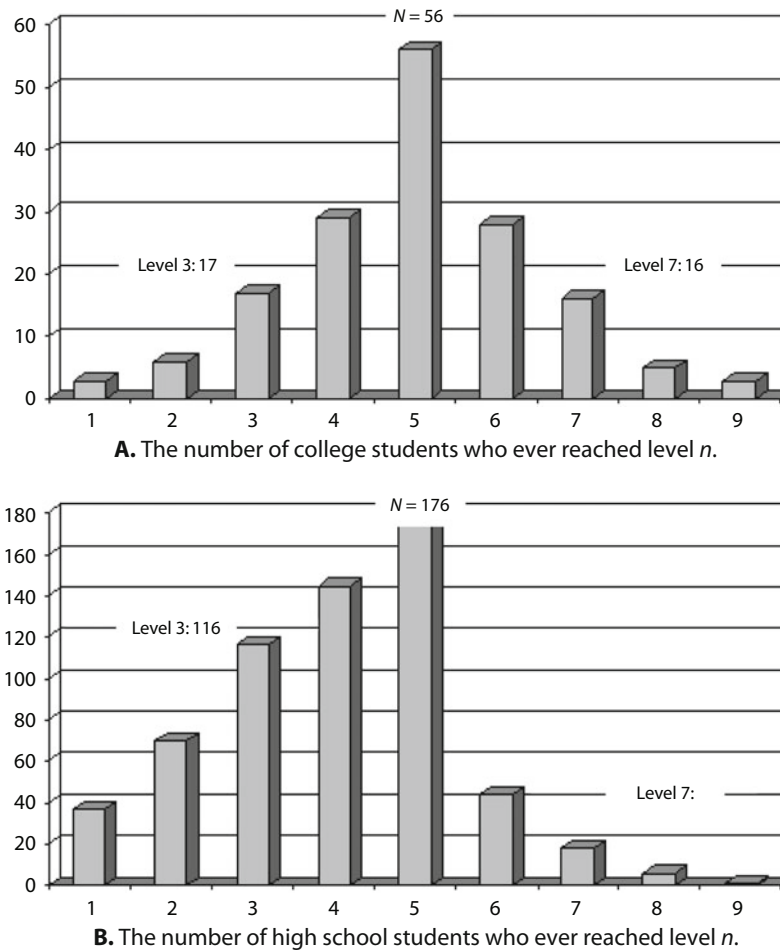


Figure 2. Comparison of college and high school student performance on the Version 1 demonstration module.

styles for a particular sentence, explanation, and student level. Designing 2 styles for 10 student levels for six sentences and their explanations means designing 120 styles. Fortunately, very few demonstrations need to be designed.

Changes to the Practice Module

In the first version of the practice module, the trainees practice self-explaining by reading one or two science texts, 12 to 20 sentences in length, and self-explaining each sentence under the guidance of a pedagogical agent, who provides feedback on their explanations. iSTART recognizes explanations that are too short, irrelevant, or too similar to the target sentence, and the coach-agent asks the trainee to try again. Explanations meeting the minimum criteria are accepted, and good explanations are praised.

During the practice sessions, the experimenters noticed that the students showed signs of frustration, especially when the second practice text appeared, and that many seemed tired by the time the session ended. Evidently the practice module was too intensive. Interviews with some students revealed that they could not envision themselves self-explaining every sentence of a textbook that they were reluctant to read in the first place. It was also observed that

when students were asked to self-explain each sentence, their explanations tended to focus on the sentences separately, and did not make connections to other sentences. To adapt the second version's practice module to the high school environment, two major changes were made: We modified the practice regimen, and we extended the practice experience over several weeks. Although the first version's regimen had the students explain every sentence, in order to provide concentrated practice experience, the new regimen had the students explain only preselected target sentences. This approach has the advantages of more nearly modeling the expected use of reading strategies outside the training environment, of providing more material for each of the explanations by providing an unexplained space of several sentences, and of allowing the students to complete each reading in a shorter time. Our research indicated that repeated practice is essential in helping high school students to learn and retain reading strategy knowledge and skills, so the second version of iSTART allows for several practice sessions, ideally spaced about a week apart. Since the preparation of a practice text in iSTART Version 1 is a major undertaking, and since the new approach increased the number of texts needed by an order of magnitude, this

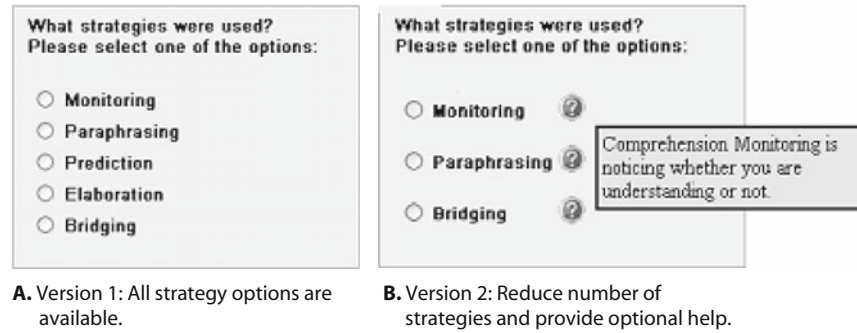


Figure 3. Demonstration dialog box comparison screen shots (Version 1 vs. Version 2).

decision led to a complete revision of our text-preparation approach, as discussed next.

Modified Self-Explanation Evaluation Algorithm

Version 2 aims to be more efficient for the experimenters, allowing experiments to be set up and evaluated with less time and effort. Some improvements in this area were also made with an eye to transitioning iSTART to a system that could be managed by a classroom teacher. The improvements discussed here include a highly expedited system for introducing new texts to be self-explained, and a computerized testing system.

The second version's practice module is a vast improvement, from the experimenters' point of view. Although the system that delivers practice to the trainee is independent of the text, each text in Version 1 required extensive preparation in order to be delivered. The preparation enables iSTART to provide feedback on the free-form explanations generated by the trainees. The feedback engine in Version 1 went through two iterations. Initially, texts required the manual identification of "important words" in each sentence, the creation of a dictionary of words associated with each of the important words, the manual

rating of the importance of the sentence in the text, and a manual determination of a minimum-length criterion for trainee explanations of the sentence. This system worked by counting partial matches of words in the trainee's explanations to words on the various lists. Then, an improved system added more work: (1) a causal analysis of the relationships among the sentences in the texts, to identify related prior sentences, and (2) the collection of terms used by participants in a preliminary experiment with the text, to capture the real-world knowledge students might use in understanding the reading. This system added the calculation of latent semantic analysis (Landauer, Foltz, & Laham, 1998; Landauer, McNamara, Dennis, & Kintsch, 2007) cosines between the student explanations and various collections of words. The creation of the dictionary, the causal analysis, and the preliminary experiment took well over a person-week of preparation for each of the two practice texts used in Version 1.

Since the second version of iSTART introduced extended practice sessions and also aimed to include a variety of genres, many new texts had to be added to the system's library. However, it would have been impossible to add them in a reasonable time, because of the extensive

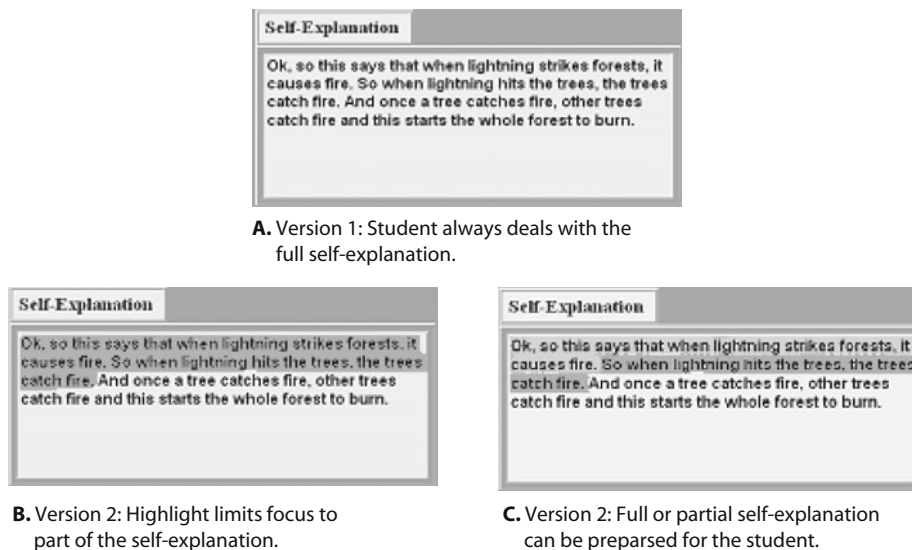


Figure 4. Demonstration self-explanation box comparison screen shots (Version 1 vs. Version 2).

preparation required for each text. Consequently, we devised a new feedback algorithm that eliminated virtually all of the human effort required to prepare a text. In brief, we identified content words automatically, used the text itself instead of creating a dictionary, mechanically chose related sentences instead of using causal analysis, and did without the preliminary experiment. In addition, we created a tool for adding texts to our repertoire that automates everything except the occasional correction of the pronunciation of words by the pedagogical agents. The tool allows nonexperts (i.e., teachers) to upload texts for their students to use. The accuracy of the new system in evaluating self-explanations appears to be as good as or better than the original labor-intensive versions (McNamara, Boonthum, Levinstein, & Millis, 2007).

Computerization of Testing

The original SERT experiments were supported by pre- and posttests consisting of multiple-choice questions as well as short-answer comprehension tests and surveys. When iSTART was first administered, it continued to use the same paper tests. The paper tests were difficult for the experimenters, because students had no standard identifiers; they might identify themselves on different instruments by given names, nicknames, and even the last names of different parents. First to be computerized was the battery of multiple-choice tests. We devised a system that provided for the creation and scheduling of tests for the experiments. Different experiments can have different sets of test questions selected from a question bank.

Subsequently, the short-answer comprehension tests were computerized. These timed tests consist of relatively lengthy readings with multiple-choice and short-answer questions. Computerization meant that all versions of the students' answers, as well as the time spent on the reading and question pages, could be stored in a database. In addition, a noncoached version of the practice module was created that acted as a self-explanation test.

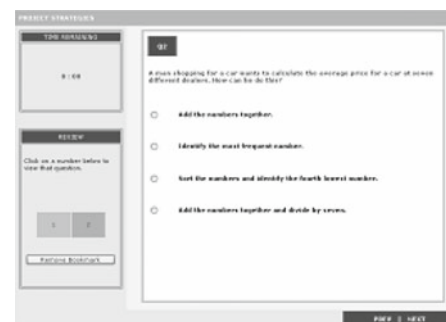
In all of our tests, after the test module is downloaded to the trainee's computer, all interaction with the server is transparent to the trainee. As the test progresses, student responses are sent to the database in the background, so that network difficulties do not interfere with the trainee's experience. We are developing reliable communication systems that will ensure that the test data eventually reach the server despite the problems caused by wireless communication, network hardware problems, and malware on high school networks. Our design also aims to preserve the navigational advantages of a paper test for the trainee (see Figure 5). In addition to the usual "previous/next" style of navigation, the multiple-choice test module allows students to jump directly to any question and to "bookmark" questions they want to return to. Similarly, the comprehension test permits the student to jump to any page in the reading and back to the question sheet.

Reactions to iSTART (Version 2)

In general, it appears that students who interact with iSTART are satisfied with the second version, as evidenced by unpublished data from a study reported by



A. An instruction page prior to the test.



B. A question with choices (no text or image). A bookmark can be set for each question.



C. A question with a passage.



D. A question with an image.

Figure 5. Screen shots of computerized multiple-choice test delivery.

McNamara et al. (2006), in which 67 high school students from four classes interacted with the program. Results from the survey data are shown in Table 3 and are based on a rating scale of 1 to 4 (low to high). These results suggest that students were motivated ($M = 2.8$, $SD = 0.5$) while working with the program, which they tended to enjoy ($M = 2.8$, $SD = 0.5$), and found it minimally difficult ($M = .30$, $SD = .60$). Although agent utility was not addressed in the creation of iSTART (Version 2), it was addressed after the second version was created, because character ratings and likeability in the original program may have been affected by challenges that students were having with other aspects of the original version.

Another question is whether students were focusing on the appropriate characters and boxes after the modifications. Eyetracking data from another unpublished iSTART (Version 2) study (Louwerse, Graesser, McNamara, Bell, & Lu, 2007), involving 7 college students interacting with iSTART for an average of 29.5 min, indicated that students looked at the agents and corresponding text balloons at the appropriate times. In spite of the fact that Dr. Julie and Mike received lower likeability scores than the other agents, according to postexperiment survey results, results from Louwerse et al. showed no significant differences between fixations on either the agents or text balloons.

Discussion

In developing this tutoring system toward what we hope will become widespread classroom use, we found ourselves dealing with three conflicting motivations. As pedagogues, we found ourselves wanting to change the system prematurely when perceived problems or attractive approaches suggested modifications that might improve its effectiveness. As administrators, we were forced by the increasing scope of the project to consider measures that would use human resources more efficiently, even when we feared they might have a cost in effectiveness. As experimenters, we were reluctant to make changes that would render future data sets incomparable with those of the past, or that could not be experimentally justified. In McNamara et al. (2004), we reported on the design and

preliminary success of the first version of iSTART, which transformed a human-delivered intervention (SERT) into a series of automated, Web-delivered modules. That article emphasized the efficiency of a Web-delivered intervention over a human-delivered one, as well as the pedagogical changes these modalities required. In this article, we confirm success of that version, but note that with continued experimentation in high school settings, we have discovered new pedagogical and administrative problems; we report on the redesign we have made to overcome them. In general, the pedagogical improvements do not increase the administrative burden, and the improved administrative efficiency has not reduced the pedagogical effectiveness, at least according to preliminary observations. Future experimentation will reveal whether these initial indications are valid.

AUTHOR NOTE

We thank the members of the ODU iSTART research lab and the Memphis CSEP lab who helped in developing new modules and improving existing ones. This project was funded by NSF IERI Grant REC-0089271 and IES Grant R305G040046 to the last author. Correspondence concerning this article should be addressed to I. B. Levinstein, Department of Computer Science, Old Dominion University, 4700 Elkhorn Avenue, Suite 3300, Norfolk, VA 23529 (e-mail: ibl@cs.odu.edu).

PREFERENCES

- 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.-H., & LAVANCHER, C. (1994). Eliciting self-explanations improves understanding. *Cognitive Science*, *18*, 439-477.
- LANDAUER, T. K., FOLTZ, P. W., & LAHAM, D. (1998). An introduction to latent semantic analysis. *Discourse Processes*, *25*, 259-284.
- LANDAUER, T. K., MCNAMARA, D. S., DENNIS, S., & KINTSCH, W. (Eds.). (2007). *Handbook of latent semantic analysis*. Mahwah, NJ: Erlbaum.
- LOUWERSE, M. M., GRAESSER, A. C., MCNAMARA, D. S., BELL, C., & LU, S. (2007). *Pay attention to embodied conversational agents*. Manuscript submitted for publication.
- MAGLIANO, J. P., TODARO, S., MILLIS, K., WIEMER-HASTINGS, K., KIM, H. J., & MCNAMARA, D. S. (2005). Changes in reading strategies as a function of reading training: A comparison of live and computerized training. *Journal of Educational Computing Research*, *32*, 185-208.
- MCNAMARA, D. S. (2004a). Aprender del texto: Efectos de la estructura textual y las estrategias del lector. *Revista Signos*, *37*, 19-30.
- MCNAMARA, D. S. (2004b). SERT: Self-explanation reading training. *Discourse Processes*, *38*, 1-30.
- MCNAMARA, D. S., BOONTHUM, C., LEVINSTEIN, I. B., & MILLIS, K. (2007). Using LSA and word-based measures to assess self-explanations in iSTART. In T. K. Landauer, D. S. McNamara, S. Dennis, & W. Kintsch (Eds.), *Handbook of latent semantic analysis* (pp. 227-241). Mahwah, NJ: Erlbaum.
- MCNAMARA, D. S., & THE CSEP LAB (2005). Promoting active reading strategies to improve undergraduate students' understanding of science. Annual project reports submitted to the National Science Foundation, IERI.
- MCNAMARA, D. S., LEVINSTEIN, I. B., & BOONTHUM, C. (2004). iSTART: Interactive strategy training for active reading and thinking. *Behavior Research Methods, Instruments, & Computers*, *36*, 222-233.
- MCNAMARA, D. S., O'REILLY, T. P., BEST, R. M., & OZURU, Y. (2006). Improving adolescent students' reading comprehension with iSTART. *Journal of Educational Computing Research*, *34*, 147-171.
- MCNAMARA, D. S., & SCOTT, J. L. (1999). Training reading strategies. In *Proceedings of the 21st Annual Conference of the Cognitive Science Society*. Mahwah, NJ: Erlbaum.

Table 3
Mean Ratings (1 = Low to 4 = High) of
Student Motivation, Enjoyment, and Difficulty

Topic	<i>M</i>	<i>SD</i>
Student Motivation	2.8	0.5
Teacher A	2.7	0.4
Teacher B	2.8	0.3
Teacher C	2.8	0.3
Teacher D	2.6	1.0
Student Enjoyment	2.3	0.6
Teacher A	2.3	0.4
Teacher B	2.2	0.5
Teacher C	2.5	0.5
Teacher D	2.5	1.0
Student Difficulties	0.3	0.6
Teacher A	0.2	0.0
Teacher B	0.2	0.6
Teacher C	0.1	0.3
Teacher D	0.7	0.9

- O'REILLY, T. P., BEST, R., & MCNAMARA, D. S. (2004). Self-explanation reading training: Effects for low-knowledge readers. In K. Forbus, D. Gentner, & T. Regier (Eds.), *Proceedings of the 26th Annual Conference of the Cognitive Science Society* (pp. 1053-1058). Mahwah, NJ: Erlbaum.
- O'REILLY, T. P., SINCLAIR, G. P., & MCNAMARA, D. S. (2004a). iSTART: A Web-based reading strategy intervention that improves students' science comprehension. In Kinshuk, D. G. Sampson, & P. Isaiás (Eds.), *Proceedings of the IADIS International Conference on Cognition and Exploratory Learning in the Digital Age: CELDA 2004* (pp. 173-180). Lisbon, Portugal: IADIS Press.
- O'REILLY, T. P., SINCLAIR, G. P., & MCNAMARA, D. S. (2004b). Reading strategy training: Automated versus live. In K. Forbus, D. Gentner, & T. Regier (Eds.), *Proceedings of the 26th Annual Conference of the Cognitive Science Society* (pp. 1059-1064). Mahwah, NJ: Erlbaum.

NOTE

1. iSTART has been developed by conducting classroom experiments, with the goal in mind to create a version for use by teachers and students. However, the current version has not been prepared for commercial use and thus is not available in a form that can be easily incorporated into a classroom by a teacher. Although Web based, the system requires that components be installed on users' computers and that users be registered with a central iSTART server for purposes of data collection, curriculum delivery, and the evaluation of student explanations. We are currently working to make the system more usable by teachers, without the assistance of experimenters.

(Manuscript received December 8, 2006;
revision accepted for publication March 15, 2007.)