Improving Understanding of Science Texts: iSTART Strategy Training vs. a Web Design Control Task

Roger S. Taylor (rtaylor@mail.psyc.memphis.edu)
Tenaha O'Reilly (t.oreilly@mail.psyc.memphis.edu)
Mike Rowe (mprowe@mail.psyc.memphis.edu)
Danielle S. McNamara (d.mcnamara@mail.psyc.memphis.edu)

Institute for Intelligent Systems
Psychology Department, University of Memphis
Memphis, TN 38152 USA

Abstract

Scientific literacy is a critical skill that requires the ability to deeply comprehend difficult scientific concepts. However, comprehending science texts is often a challenge for low-knowledge readers. This research assessed the effectiveness of an automated reading strategy training program (iSTART) to improve comprehension for low-knowledge students. Eighty-four high school students were either trained to use iSTART or they were given instruction on how to design science web pages (i.e., control). Following a several month delay after training it was found that the students in the iSTART condition comprehended more from a science text than students in the control condition.

Introduction

The impact of science literacy on our economy and society is significant, and has been widely documented (e.g. AAAS, 1993; NRC, 1996). The recent report by the National Academy of Sciences (NAS, 2005) describes how American leadership in the areas of science and technology has eroded, and lays out the serious economic problems that will result if this trend is not reversed. A primary mode of pre-collegiate science instruction involves the reading of expository science texts. Given that the intended purpose of such texts is the introduction of new and unfamiliar concepts, it is not surprising that many students have difficulty reading such materials (Bowden, 1999; Snow, 2002). This problem is made more difficult because students are often not equipped with the knowledge to succeed in their courses (Snow, 2002) and furthermore, students do not typically use good reading strategies (Cox, 1997; Garner, 1990). Therefore, one promising approach is to provide students with training in the use of more effective reading strategies, which they can employ when reading expository science texts. Indeed, empirical studies have demonstrated that interventions focused on teaching reading strategies have been successful in improving reading comprehension (e.g., Pressley et al., 1992; Rosenshine & Meister, 1994; Rosenshine, Meister, & Chapman, 1996).

Recently, attention has been given to a reading intervention called Self-Explanation (Chi, De Leeuw, Chiu, & LaVancher, 1994). Self-explanation is a process in which a learner explains aloud the meaning of challenging material, such as science texts. Chi et al.'s research has demonstrated large gains in learning for students when they were prompted to self-explain. McNamara and colleagues have built upon this research by providing learners with scaffolded training in metacognitive reading strategies (McNamara, 2004). In the training program, called self-explanation reading training (SERT), students learn to use metacognitive reading strategies (e.g., comprehension monitoring, paraphrasing, elaboration, and bridging inferences). Empirical studies have indicated that SERT improves comprehension, particularly for low-knowledge readers (e.g., McNamara, 2004; O'Reilly, Best, & McNamara, 2004).

Interactive Strategy Training for Active Reading and Thinking: iSTART

While previous investigations of SERT were encouraging, there are also two limitations of its future implementation on a larger scale. First, the high cost of human tutors makes widespread adoption of the program problematic, particularly in economically distressed areas. Second, the structure of the SERT program does not allow for dynamic instruction that is tailored to the specific needs of the learner. Consequently, McNamara and colleagues developed an automatized version of SERT training called iSTART (Interactive Strategy Training for Active Reading and Thinking; see McNamara et al., 2004). Training with iSTART occurs in three phases (Introduction, Demonstration, and Practice) and takes approximately 2 1/2 hours. The delivery of information and feedback to students is accomplished via animated pedagogical agents. The content of the feedback is determined based on the quality of the students' self-explanations, as assessed via a set of linguistic algorithms (for details, see McNamara et al., 2004; McNamara, Boonthum, Levinstein, & Millis, in press).

The first phase of training is strategy *Introduction*. This phase includes definitions and examples of the process of self-explanation as well as five reading strategies (comprehension monitoring, paraphrasing, prediction, bridging inferences, and elaboration). After each strategy is presented, students are asked to answer four multiple-choice questions and are then provided with immediate feedback by the program. The second phase of training is strategy *Demonstration*. In this phase, pedagogical agents model the use of the reading strategies while they are self-explaining a science text. The student is asked to identify the strategies used by the pedagogical agent during self-explanation, and is provided with feedback. The third phase of training is strategy *Practice*. In this phase, students read two short science passages and are asked to apply the newly learned strategies while typing self-explanations of the sentences in the texts. As they proceed through the text, several algorithms are employed to evaluate the quality (e.g., self-explanation length and the number and type of content words) of the generated self-explanations (McNamara et al., 2004). Based upon this analysis, scaffolded feedback is provided. For instance, a student who gives an "impoverished" self-explanation that does not go beyond the content of the text's sentence is prompted by the agent to provide more information.

Current Study

Previous investigations have revealed that iSTART improves comprehension as compared to normal reading controls (O'Reilly, Sinclair, & McNamara, 2004a). However, many students lack the knowledge in order to succeed in their courses (Snow, 2002). Therefore, a principle aim of this study was to determine whether iSTART can help low-knowledge readers learn more from typical high school level science texts. While prior investigations with human trained self-explanation strategy training (SERT) have shown benefits for low-knowledge readers (e.g., McNamara et al., 2004), it was not clear whether these results would transfer to iSTART training. In particular, we were interested in whether the effect of the strategy training would hold under challenging but real-world conditions. Therefore, we tested the effectiveness of the training in live classrooms during regular class time. Furthermore, we also assessed the impact of the training after an extended delay so that we could determine whether the training was flexible enough to persist over time.

In order to provide a challenging comparison with iSTART, students in a control condition were taught to design web pages containing science-related information. The web design task was developed to engage students with a computerized activity, while at the same time being exposed to the same scientific content as the iSTART group. This kind of control condition accounts for effects of novelty pertaining to computer use as well as exposure to the scientific information involved in iSTART training.

In the current study, we investigated the efficiency with which the iSTART program can be used in the classroom to teach high school students reading strategies. A major experimental goal was to assess potential improvements of students' science text comprehension as a function of iSTART training, particularly for the students most in need of remediation – those who possess the least amount of scientific domain knowledge. We predicted that, because the iSTART system is adaptive to the specific level of the student, low-knowledge students who were trained in iSTART would learn more from science texts than those in the control condition.

Method

Participants

The initial sample consisted of 446 students in 10 physical science classes (9th grade) and 10 biology classes (10th grade) from a suburban Tennessee high school. Of these students, 153 completed all of the training, pretest, and posttest assessments under investigation. A median split was performed on prior science domain knowledge and the resulting 82 low-knowledge students were included as the participants for this study.

Assessment Materials

Reading Ability General reading skill was measured using a modified version of the standardized Gates-MacGinitie reading skill test for grades 7-9 that consisted of 48 multiple-choice questions which assess student comprehension on several short text passages.

Prior Science Knowledge Prior general science knowledge was measured with a 20-item, four-alternative, multiple-choice test. The test covered several areas including biology, chemistry, earth science, research methods and mathematics. Questions were selected from high school science tests collected from several states (i.e., Colorado, Georgia, Kentucky, Tennessee, and Virginia).

Pretest Science Passage Comprehension Two passages were presented to the students to assess pre-training science text comprehension. The first pretest passage was a 449-word passage on petroleum that described the refining process of heating petroleum to different temperatures, thereby allowing for the separation and collection of methane versus diesel fuel and other similar products. The 31-sentence passage had a Flesch Reading Ease of 39.7 and a Flesch-Kincaid Grade Level of 11.2. The second pretest passage was a 322-word passage on the carbon cycle that described how carbon passes through the food supply of the ecosystem. The 17-sentence passage had a Flesch Reading Ease of 50.5 and a Flesch-Kincaid Grade Level of 10.6. There were eight open-ended questions presented for each passage. The answers to four of the questions could be found within a single sentence of the passage, and they are referred to as *text-based* questions. The answers to the remaining four questions required the reader to combine information contained in two or more sentences of the passage, and are referred to as *bridging-inference* questions.

Posttest Science Passage Comprehension Two passages were presented to the students to assess post-training science text comprehension. The first posttest passage was a 307-word passage on the most common types of medical problems involving the heart. The 21-sentence passage had a Flesch Reading Ease of 55.9 and a Flesch-Kincaid Grade Level of 9.0. The second posttest passage was a 477-word passage on the origins of the universe, which described theories related to the Big Bang. The 31-sentence passage had a Flesch Reading Ease of 39.1 and a Flesch-Kincaid Grade Level of 11.5. There were eight open-ended questions presented for each passage. As with the pretest passages, there were four text-based questions and four bridging inference questions.

Design and Procedure

The experiment used a pretest, intervention, posttest design with a between-subjects manipulation comparing iSTART strategy training (experimental condition) with web design training (control condition). Using matched-samples assignment (based on reading skill and prior knowledge), half of the students in each classroom were assigned to the iSTART condition, and half were assigned to the control condition.

The experiment consisted of three phases: pretest, training, and posttest, with the pretest and posttest phases being identical for both the iSTART and control conditions. During the *pretest*, students in the iSTART and control group were administered the pretest measures in the following order and time frame: prior science knowledge (10 min), Gates-MacGinitie reading measure (Form K, 15 min), and two science passages on petroleum and the carbon cycle, along with the sets of comprehension questions (15 min).

Based on pretest scores on reading skill and prior knowledge, the experimenters used a matched-samples assignment to randomly assign half the students to the iSTART condition and the other half to the control condition. The *training* phase (described in detail below) lasted for four days.

In order to better measure the robustness of any potential effects of training, there was a planned "delay" between the training phase and the posttest phase. Logistical classroom constraints required that for approximately half of the participants in both conditions, the training phase occurred two months before the posttest phase, while for the remaining students, the training phase occurred five months before the posttest phase.

During the *posttest*, students in both groups were administered the posttest measures in the following order and time frame: two science passages on heart disease and the universe, along with corresponding sets of comprehension questions (15 min).

iSTART Strategy Training iSTART training was facilitated by University of Memphis researchers and was completed over four consecutive days. Students worked through the iSTART modules in a sequential order – Introduction, Demonstration and Practice. Training required a total of 2.5 to 3 hours, and occurred over four class periods. During the introduction phase, students were provided with information on five reading strategies (i.e., comprehension monitoring, paraphrasing, prediction, bridging inferences, and elaboration). During the demonstration phase, pedagogical agents modeled the previously presented reading strategies. Lastly, during the practice phase, students typed their self-explanations for a text on thunderstorms followed by a text on coal. Target sentences were self-explained in a sequential order, on a sentence-by-sentence basis. Self-explanations appeared in part of the screen called the self-explanation box. After completing each self-explanation, the participant was required to submit their self-explanation to be evaluated by the pedagogical agent. The student was then given feedback (e.g., the agent requested more details to be added, or offered a hint that the student should add more details next time). Once the final protocol for a given target sentence had been accepted, the next sentence was displayed.

Web Design Training In order to provide a fair comparison between conditions, students in the web design condition covered the same science content as those in the iSTART condition. Across four days, control students were trained in the creation of web pages, which was facilitated by University of Memphis researchers. The first day was spent learning and practicing the tags needed for web page creation. Template pages were provided as a starting point for each lesson. Led by an instructor, students opened the templates and saved them under a new name as the foundation for their web page. Using their new web pages, students were instructed in the proper use of the hypertext markup tags (e.g., <p></p>, <a>, and
). Each student was instructed to follow along and use the tags in their own page. After all the tags were demonstrated, the students were given tasks to practice the tags' proper use.

On the second day students were instructed to create a web page to answer science questions using the tags they learned the previous day. Study sheets containing all the tags and examples of their use were provided along with the questions the students were to answer. The instructor answered a sample question first by again showing students how to create a web page from the template. This question was then used as a title for the web page. Answers to the questions could be found on various text-based science web pages created by the experimenters linked to a main page saved on each computer. The instructor opened a science web page that answered the question and copied the content to the body of the web page. The name of the page from which the content was taken was also copied and placed beneath the answer. After each student completed the example, they were told to answer the remaining questions in the same way.

Day three and day four followed the same procedure. However, instructors no longer provided a demonstration to begin each lesson. Instead, instructors provided support and helped troubleshoot students' problems as needed. The students were provided with study sheets along with a list of science questions they were to answer that day, as well as a selection of science web pages from which they were to answer each question.

Results

Pretest Individual Difference Scores

There were no significant differences in pretest prior science knowledge scores between the iSTART condition (M = 0.53, SD = 0.16) and the control condition (M = 0.52, SD = 0.17), $t(71) = 0.39$, $p = .70$ (2-tailed). There were also no significant differences in pretest reading ability scores between the iSTART condition (M = 20.74, SD = 8.30) and the control condition (M = 20.98, SD = 7.41), $t(147) = 0.18$, $p = .86$ (2-tailed). Therefore, any effects of condition cannot be accounted for by any pre-training differences in prior science knowledge or reading ability.

Text-Based Questions

The proportion of correctly answered text-based questions (n = 8) is depicted in Table 1. Analyses were performed to ensure that there were no significant differences between the two levels of delay (i.e., two month vs. five month gap between the training and posttest phases). Since no significant effects were found in this analysis, the two levels of the delay were collapsed across the remaining analyses.

Table 1. iSTART and control condition values for text-based (TB) and bridging inference questions (Br).

Condition	Pretest	Posttest	p value	Cohen's d
iSTART (TB)	.306 (.174)	.522 (.206)	< .001	1.13
Control (TB)	.329 (.184)	.448 (.233)	.014	0.57
iSTART (Br)	.091 (.075)	.278 (.171)	< .001	1.42
Control (Br)	.127 (.094)	.258 (.194)	.001	0.86

Performance gains for students were examined via a 2 (iSTART, Control) X 2 (Pretest, Posttest) Mixed Model ANOVA. There were significant gains in performance from pretest to posttest for all participants, $F(1, 80) = 35.12$, $p < .001$. The main effect of Condition was not significant, $F(1, 80) = 0.54$, $p = .463$. However, the Test by Condition interaction was marginally significant, $F(1,80) = 2.92$, $p = .091$ (see Figure 1).

Further analyses were conducted to determine whether the magnitude of the learning gains were different as a function of condition. As predicted, an independent sample t-test revealed that the students in the iSTART condition had a significantly higher gain in learning in comparison to those in the control condition, $t(80) = 1.71$, $p = .046$ (one-tailed), Cohen's $d = 0.38$ (see Table 1). These results demonstrate how, in comparison to the control condition, the iSTART strategy training leads to significantly greater performance on text-based reading comprehension questions.

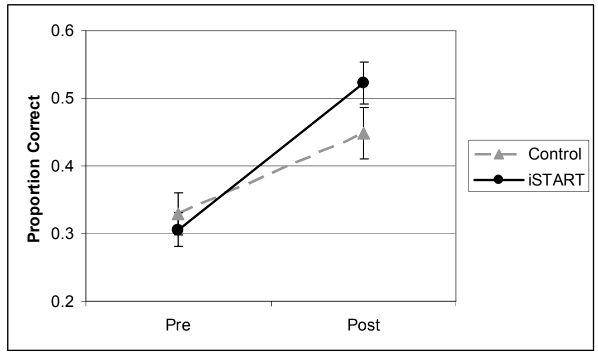

Figure 1. Proportion of correct textbase questions as a function of training condition (iSTART vs. Control).

Additional analyses (i.e., paired samples t-tests) were performed to examine pretest/posttest changes for both conditions. As predicted, there was a significant increase in performance for students in the iSTART condition, $t(48) = 6.21$, $p < .001$, Cohen's $d = 1.13$. Similarly, there was a significant increase in performance for students in the control condition, $t(32) = 2.61$, $p = .014$, Cohen's $d = 0.57$ (see Table 1). Thus, both conditions significantly improved from pretest to posttest – with a moderate increase for the control condition ($d = 0.57$) and a large increase ($d = 1.13$) for the iSTART condition.

Bridging Inference Questions

The proportion of correctly answered bridging inference questions (n = 8) was used to assess performance on the bridging inference questions. Analyses were performed to ensure that there were no significant differences between the two levels of delay (i.e., two month vs. five month gap between the training and posttest phases). Since no significant effects were found in this analysis, the two levels of the delay were collapsed across the remaining analyses.

Performance gains for students were examined via a 2 (iSTART, Control) X 2 (Pretest, Posttest) Mixed Model ANOVA. There were significant gains in performance from pretest to posttest for all participants, $F(1, 80) = 61.92$, $p < .001$. The main effect of Condition was not significant, $F(1, 80) = 0.11$, $p = .740$. The Test by Condition interaction was not significant, $F(1,80) = 1.97$, $p = .165$.

Further analyses were conducted to determine whether the magnitude of the learning gains was different as a function of conditions. As predicted, an independent sample t-test revealed that the students in the iSTART condition had a marginally higher gain in learning in comparison to those in the control condition, $t(80) = 1.40$, $p = .083$ (one-tailed), Cohen's $d = 0.30$ (see Table 1). These results provide support, albeit somewhat weak, that the iSTART strategy training can lead to better performance on bridging inference questions in comparison to the control condition.

Additional analyses (i.e., paired samples t-tests) were performed to examine pretest/posttest changes for both conditions. There was a significant increase in performance for students in the iSTART condition, $t(48) = 8.55$, $p < .001$, Cohen's $d = 1.42$. Similarly, there was a significant increase in performance for students in the control condition, $t(32) = 3.53$, $p = .001$, Cohen's $d = 0.86$ (see Table 1). In sum, in addition to the statistical significance, in terms of effect size, both the control condition ($d = 0.86$) and the iSTART condition ($d = 1.42$) had large increases from pretest to posttest.

Discussion

The principal aim of this study was to examine the impact of an adaptive computer-based reading strategy tutoring system (iSTART) on high school students' comprehension of expository science texts. The focus was to assess the potential facilitation of comprehension for the students who need it the most –low-knowledge readers. In addition, we sought to test the system under ecologically valid conditions by assessing it in high school classrooms. Our prediction was that low-knowledge readers trained with iSTART would have significantly greater gains in reading comprehension performance, in comparison to those in the control condition.

The results of this study provided some support for our predictions – participants in the iSTART condition outperformed students in the control condition on text-based questions. While participants in the iSTART condition scored higher than the control on bridging inference questions, this effect was only marginally significant.

These results echo earlier findings that have shown that iSTART can improve comprehension (O'Reilly, et al., 2004a), and that SERT based training improves comprehension for low-knowledge readers (McNamara, 2004), primarily on text-based questions. We are certainly not surprised to see a larger effect on text-based questions. Enhanced effects of reading strategy training (using SERT or iSTART) for low-knowledge (or less-skilled readers) on text-based questions have been observed in five previous studies (Magliano, et al., 2005; McNamara, 2004; O'Reilly, Sinclair, & McNamara, 2004a,b; McNamara, O'Reilly, Best, & Ozuru, in press). This result indicates that reading strategy training allows low ability readers to better understand the basic ideas in the text – certainly a worthwhile goal. We're hopeful that additional, extended training for low ability readers will help them to go beyond textbase level comprehension and learn to understand text more deeply.

This is the first study to show that iSTART is effective in high school classrooms, and in particular, it is the first study to demonstrate that iSTART helps improve comprehension for *low-knowledge readers*. Also noteworthy is that the beneficial effects of training persisted over an extended period of time and were evident months later. Lastly, this experiment employed a challenging contrast to the iSTART intervention – the active and engaging process of designing science web pages as a control condition.

The control condition in this study was quite stringent in that it controlled for time on task, the engagement level of the task, the task being computer based, and the scientific information read by the students during the task. Given that this experiment occurred over a period of several months, the improvement in student reading ability and knowledge levels may have resulted from their normal academic experiences. However, we observed significantly higher gains for students in the iSTART condition than those in the control condition, which indicates that iSTART effectively improved students' ability to understand science text over and above their normal academic experiences.

In conclusion, the growing recognition of the national importance of science literacy highlights the need to provide students with the ability to learn from expository science texts. The results presented here demonstrate how the use of powerful educational tools such as the iSTART instructional system can help make this possible.

Acknowledgements

We would like to thank Grant Sinclair, Pavan Pillarisetti, and Chutima Boonthum. This project was supported by the NSF (REC-0241144) and IES (R305G040046). Any opinions, findings and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the NSF or IES.

References

American Association for the Advancement of Science (AAAS), Project 2061. (1993). *Benchmarks for science literacy*. New York: Oxford University Press.

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

Chi, M. T. H., De Leeuw, N., Chiu, M., & LaVancher, C. (1994). Eliciting self-explanations improves understanding. *Cognitive Science, 18*, 439-477.

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.

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

Magliano, J. P., Todaro, S., Millis, K. 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. (2004). SERT: Self-Explanation Reading Training. *Discourse Processes, 38*, 1-30.

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

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., Best, R., & Ozuru, Y. (in press). Improving adolescent students' reading comprehension with iSTART. *Journal of Educational Computing Research*.

National Academies of Sciences (NAS). (2005). *Rising above the gathering storm: Energizing and employing America for a brighter economic future*. Washington, DC: National Academy Press.

National Research Council (NRC). (1996). *National science education standards*. Washington, DC: National Academy Press.

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 Meeting of the Cognitive Science Society* (pp. 1053-1058). Mahwah, NJ: Erlbaum

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

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

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

Rosenshine, B., & Meister, C. (1994). Reciprocal teaching: A review of the research. *Review of Educational Research, 64*, 479-530.

Rosenshine, B., Meister, C., & Chapman, S. (1996). Teaching students to generate questions: A review of the intervention studies. *Review of Educational Research, 66*, 181-221.

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