Interactive Paraphrase Training:
The Development and Testing of an iSTART Module

Danielle S. McNAMARA\textsuperscript{a,1}, Chutima BOONTHUM\textsuperscript{b}, Christopher A. KURBY\textsuperscript{a}, Joe MAGLIANO\textsuperscript{c}, Srinivasa PILLARISSETTI\textsuperscript{a}, Cedrick BELLISSENS\textsuperscript{a}

\textsuperscript{a}University of Memphis
\textsuperscript{b}Hampton University
\textsuperscript{c}Northern Illinois University

Abstract. Comprehension of science texts is challenging, particularly when the reader lacks the skills or knowledge necessary to fill in conceptual gaps in the text content. The iSTART system was developed to help readers learn and practice reading strategies to improve their ability to comprehend challenging text. This study describes a new iSTART module recently developed and tested, called Interactive Paraphrasing (IP), in which students are interactively and adaptively taught how to paraphrase sentences. We compared the effects of iSTART to iSTART with IP (IP-iSTART) with high school students on their strategy use and ability to comprehend text. IP-iSTART increased skilled readers’ self-explanation quality, improved their ability to answer online comprehension questions, and increased their use of paraphrases after training. Less skilled readers benefited most in self-explanation quality from the original version of iSTART. Results are discussed in terms of tailoring reading strategy training to the needs of the reader.

Keywords. Paraphrase Evaluation, Feedback, Reading Strategy, Self-Explanation

Introduction

Many contend that the future of affordable, high-quality education lies in harnessing the potentials of computer technologies. While implementing computer technologies in schools has had both failings and challenges \cite{1}, significant progress in the quality of education to some extent depends on our ability to leverage the many advantages of computer technologies. Specifically relevant to this study are computer technologies with interactive tutoring that can mimic very closely the best features of human one-on-one tutoring. Computer technologies enable adaptive, one-on-one tutoring to virtually all students in the classroom.

The need for more effective instruction is evidenced by the performance of our nation’s students on national and international reading tests. Students in our educational system need more help in learning how to read, and moreover learning how to \textit{read to understand} and to \textit{reason with} the material they read. Students in the United States typically score lower on measures of reading comprehension as compared to students in other countries \cite{2} and as many as 26\% of 8th grade students cannot read at

\footnote{1 Corresponding Author}
the basic level [3]. In other words, these 8th grade students do not understand what they read. The ability to read and to understand what is written is critical to success in our society as well as success in our educational system. One consequence of low reading ability is the inability to understand the material conveyed in textbooks. This deficit has the greatest consequences on the most challenging texts - science texts. Science texts are replete with technical terms, often describing complex mechanisms with multiple components, relations between components, and dynamic processes [4]. Students also tend to lack domain knowledge in science [5]. Thus, reading problems can become most apparent when students are faced with challenging science texts (e.g., [6]).

To address this need, iSTART (Interactive Strategy Training for Active Reading and Thinking) was developed to help high school and college students learn and practice reading comprehension strategies. Its design is based on a classroom intervention called Self-Explanation Reading Training (SERT; [7]) that combines self-explanation [8] with reading comprehension strategies [9]. Students learn to self-explain using five reading strategies: monitoring comprehension, paraphrasing, making bridging inferences between the current sentence and prior text, making predictions, and elaborating the text with links to what the reader already knows. There are three phases to training (Figure 1). The first phase involves defining and giving examples of the strategies through a simulated conversation between a teacher and student animated agents. The second phase involves animated agents modeling the use of the strategies coupled with interactive activities. Then in the final phase, users practice using the strategies while reading a scientific text. During this phase, an animated agent provides scaffolding and feedback to students based on computational algorithms that evaluate the quality of student input [11], [12].

Figure 1. iSTART Trainer
1. Interactive Paraphrase Training

This study examines the effects of the first additional module that we have developed and tested, Interactive Paraphrasing (IP). IP was developed to provide students with extra practice in paraphrasing sentences. Paraphrasing is an important component of comprehension. Paraphrasing alone leads only to a shallow, text-based understanding of material. However, paraphrasing lays the foundation for the reader to be able to make inferences for the construction of a deeper understanding of the material. So, in combination with other strategies, paraphrasing is quite effective and important. The inclusion of these strategies in a reading strategy intervention is motivated by theories of reading comprehension (e.g., [13]) as well as reading strategy research. Theories of reading comprehension propose that there are multiple levels of comprehension. At a minimum, there are three levels of comprehension: surface structure (reader’s memory for words and syntax); textbase (memory for the meaning behind the words and syntax, or at the propositional level), and situation model (understanding from knowledge-based inferences that go beyond the text).

Paraphrasing is assumed to help a reader form a more coherent textbase, and the inference strategies are assumed to support a more coherent situation model. This theoretical assumption has been supported empirically. McNamara et al. [14] found that more effective paraphrases were associated with better performance on questions that were designed to assess the reader’s textbase level of understanding. Specifically, paraphrasing led to better performance on text-based questions that required only one sentence from the text to answer. Further, the presence of inferences during self-explanation was associated with better performance on bridging-inference questions, which require understanding unstated relationships between separate ideas in the text. Because bridging-inference questions assess the understanding of relationships in the text, they are assumed to assess the reader’s situation model level of understanding.

Paraphrasing provides the starting point for effective self-explanations, and thus the ability to paraphrase has appeared to be critical to students’ gain from iSTART. The results showing that less-skilled readers tended to gain in terms of their ability to understanding at the textbase level, and correspondingly, to better phrase text (e.g., [14], [15]), prompted the development of a separate module to provide training and practice on paraphrasing. While iSTART already provides instruction on paraphrasing, it does not encourage the strategy of paraphrasing alone. During practice, it encourages students to go beyond the paraphrase and to generate inferences about the text.

One question here was whether providing an opportunity to practice paraphrasing, and providing positive feedback on successful paraphrases would help to boost the effects of iSTART. On the one hand, a benefit for IP-iSTART would be expected because the extra paraphrasing practice would help to boot-strap students’ performance during iSTART self-explanation practice. That is, once students had the chance to practice paraphrasing, then when it was time to practice self-explanation, then they could concentrate their efforts on self-explanation, rather than paraphrasing. On the other hand, we were concerned that such practice on paraphrasing would provide too much encouragement to paraphrase (which does not lend to deep understanding by itself), and potentially even confuse the students. That is, during one part of the training, there would be encouragement to paraphrase, and then in the final part (during iSTART self-explanation practice), the students would be discouraged from (only) paraphrasing.

A second question was whether the effects of IP-iSTART would depend on individual differences. We expected that the effects of IP would depend primarily on
students’ prior reading ability. We did not expect interactions with prior knowledge because knowledge primarily drives readers’ ability to make inferences. Knowledge is not particularly critical to the ability to paraphrase a sentence. By contrast, the ability to paraphrase is a component of reading skill, and thus we expected that whether students benefited from such training would be related to reading skill.

2. Method

Participants, Design, and Procedure

Thirty-five students (age 16-18; 28 female, 7 male) from a Mid-South high school participated for $80 ($10 each for the first three sessions and $50 for the last session). A female student in the IP-iSTART did not return after the first session. This resulted in a total of 18 students in the iSTART condition, and 17 in the IP-iSTART condition.

This experiment was conducted as a training condition 2 (iSTART vs. IP-iSTART) x 2 testing occasion (pretest vs. posttest) mixed design. The self-explanation tests (i.e., Birds, Bees) and RSAT (i.e., forms A and B) were paired as groups and counterbalanced across pretest and posttest. All participants were administered the same pretest measures on Day 1. Prior to Day 2, participants were randomly assigned to condition with the constraint that there were no between-condition differences on the Gates-MacGinitie, the prior knowledge measure, and the RSAT.

There were four sessions: (1) Pretests: participants were first administered a computerized version of the GMRT (20 minutes), the prior knowledge (PK) test, a demographics questionnaire, RSAT, and finally a self-explanation test. (2) Training-Part 1: participants proceeded through iSTART Introduction (with or without IP) and Demonstration; (3) Training-Part 2: participants engaged in the iSTART Practice module; and (4) Posttests: participants were administered RSAT and the second self-explanation test.

Materials

Participants were administered pretest measures and posttest measures. Pretest included the Gates-MacGinitie (level 7-9), a science prior knowledge test, a self explanation test, and the Reading Strategy Assessment Tool (RSAT).

Gates-MacGinitie Reading Test (GMRT). Reading skill was measured by the GMRT reading skill test (form L) level 7/9 [16]. The GMRT consists of 48 multiple-choice questions designed to assess student comprehension on 14 short text passages. The test was timed for 20 minutes. We were solely interested in reading comprehension and not vocabulary scores; thus, the vocabulary section of the test was not administered.

Prior knowledge. This is a 20 item, four-alternative multiple-choice, test that assesses knowledge of different science domains such as biology, and scientific methods, which we have used in our previous research [14], [17]. The scores from this test were used during the assignment of participants to condition to equate the training groups on prior knowledge.

Self-explanation tests. We used two science texts as materials for the self-explanation tests. One text was about Bee behavior (Bees) and the other was about the muscular structure of birds and how it contributes to their ability to fly (Birds). Both texts had similar features [18]: 24-26 sentences, a Flesch Reading Ease of 65, an adjacent argument overlap score between 0.6-0.7, and a Flesch-Kincaid grade level of 8.
There were eight pre-determined target sentences for each text. At each target sentence, participants were asked to type in their self-explanation.

Reading Strategy Assessment Tool (RSAT). RSAT is a computer-based instrument that assesses the use of reading strategies and comprehension processes with open-ended questions while they are reading a text [19]. Users read texts one sentence at a time on a computer and answer two types of questions at pre-selected sentences. Direct questions require readers to give reasons for why an event is happening in the context of the text. Indirect questions take the form of “What are you thinking now?” and users are instructed to answer these questions in a fashion that is consistent with thinking aloud. Answers to the direct questions are indicative of reading skills, whereas indirect questions assess strategy use prior to and after training (see [19] for more details). We used two counterbalanced forms of RSAT. Each form consisted of two texts developed to be equivalent in terms of assessing comprehension processes and skill.

iSTART vs. IP-iSTART

The iSTART condition received the version of iSTART described in [20]. Briefly, iSTART consists of three modules: introduction (defines and provides examples of self-explanation and the reading strategies), demonstration (shows the use of the strategies through the interaction), and practice (allows students to practice the newly learned strategies during reading). The self-explanations are automatically evaluated and immediate feedback is given to the students (see [11] and [20] for more details).

The IP-iSTART condition received the same version of iSTART as above but with a modification including IP within the introduction module. IP includes additional examples and practice paraphrasing in two phases: (1) an introduction providing a definition of paraphrasing and an explanation for how paraphrasing can help a reader better understand texts; and (2) paraphrasing practice where the reader paraphrases a sentence with encouragement to pay attention to the relations between concepts (Figure 1.d). There were 12 interactive examples, and the number of examples that students practiced depends on the students’ performance. Students were given two additional opportunities, in addition to the initial attempt, to improve their paraphrases to the acceptable level set by the system’s algorithms. When a paraphrase produced in the first attempt was unsatisfactory, the system provided feedback (e.g., too short). Students were asked to respond to the feedback by revising their paraphrase.

The evaluation algorithm used in IP-iSTART was similar to that used in the self-explanation phase of the iSTART system. That is, the student’s explanation is compared against four benchmarks (target sentence, immediate prior sentences, other prior sentences, and subsequent sentences) using Latent Semantic Analysis (LSA, [10]) and word matching (both literal and Soundex matching). A series of filtering criteria assesses whether the explanation is too short, too similar to the target sentence, or irrelevant. Length is assessed by a ratio of the number of words in the explanation to the number in the target sentence, taking into consideration the length criterion. Relevance is assessed from the number of matches to content words in the text (i.e., all sentences). Similarity is assessed in terms of a ratio of the sentence and explanation lengths and the number of matching important words. If the explanation is close in length to the sentence, with a high percentage of word overlap, the explanation would be deemed too similar to the target sentence. If the explanation fails any of these three criteria (Length, Relevance, and Similarity), the student would be given feedback corresponding to the problem and encouraged to revise the self-explanation. Once the explanation passes the filtering criteria, then it is evaluated in terms of its overall
quality. Formulae are based on the length of the sentence, the length of the explanation, the length criterion, the number of matches to benchmarks, and cosine distances to benchmarks. This algorithm has been developed with high reliability and validity and corresponds well to human scores showing correlations above 0.65 and d primes from 1.2 to over 2.0 depending on the particular algorithm, text, and target population (see detail in [11]).

In IP-iSTART evaluation algorithm, the threshold values that determine whether a protocol is short or too similar were modified because paraphrases are often shorter than self-explanations. Based on this algorithm, the teacher agent provided verbal feedback to the students about the quality of the paraphrase. Students were given a minimum of 6 examples to complete, and additional examples if performance was below the following criteria: (a) three acceptable paraphrases without feedback, including two out of three, consecutive targets; or (b) a total of six acceptable paraphrases without feedback; or (c) completed all 12 IP examples. Given that this was an initial test of the IP module, the criteria and paraphrase evaluation algorithms were identical for all participants.

3. Results

Correlations. The correlations among prior knowledge, GMRT, and RSAT scores, indicated moderate correlations between prior science knowledge and GMRT reading ability (r=.44, p<.01) as well as prior science knowledge and the RSAT comprehension score (r=.44, p<.05). No other correlations were significant.

Self-Explanation Quality. Participants were classified as skilled or less-skilled comprehenders based on a median-split using the GMRT (median = 0.46). The mean GMRT score for skilled readers was 0.62 and for less-skilled readers was 0.35, t(34)= 7.32, SE=.04, p < .001. There were 9 low and 8 high GMRT reading ability students in the IP-iSTART condition, and 10 low and 8 high GMRT reading ability students in the iSTART condition. We conducted a 2 (training condition: IP-iSTART vs. iSTART) x 2 (testing occasion: pretest vs. posttest) x 2 (GMRT reading ability: High vs. Low) ANOVA. The dependent variable was self-explanation quality as determined by the iSTART algorithm.

Table 1 presents the means and standard deviations for each group. Initial analyses confirmed that the groups did not significantly differ on pretest self-explanation performance (all ps > .26). Overall, self-explanation quality increased from pretest to posttest (pretest = 1.67; posttest = 1.90), F(1, 31)=5.22, MSE=.18, p < .05. There was a 3-way interaction that indicated that change from pretest to posttest depended on training condition and GMRT reading ability level, F(1, 31)= 4.31, MSE = .18, p < .05. Follow-up tests shown in Table 3 indicated that for the IP-iSTART condition, there was a marginally significant increase in self explanation quality from pretest to posttest for the high ability students, t(7)=2.02, SD= .66 , p=.083. For the iSTART condition, there was a significant pretest-posttest increase for the low ability students, t(9) = 2.44, SD=.53 , p < .05. All other separate effects were not significant (ps > .20). Thus, iSTART benefited low ability readers, but IP-iSTART benefited high ability readers in terms of its effect on self-explanation quality.
Table 1. Means and standard deviations (in parentheses) of self-explanation quality for each condition and reading ability level (GMRT).

<table>
<thead>
<tr>
<th>GMRT Reading Ability Level</th>
<th>Training Condition</th>
<th>Pretest</th>
<th>Posttest</th>
<th>Pre-Post t</th>
</tr>
</thead>
<tbody>
<tr>
<td>Low</td>
<td>IP-iSTART</td>
<td>1.71 (.39)</td>
<td>1.51 (.40)</td>
<td>-1.41</td>
</tr>
<tr>
<td></td>
<td>iSTART</td>
<td>1.58 (.32)</td>
<td>1.99 (.59)</td>
<td>2.44**</td>
</tr>
<tr>
<td>High</td>
<td>IP-iSTART</td>
<td>1.56 (.77)</td>
<td>2.03 (.46)</td>
<td>2.02*</td>
</tr>
<tr>
<td></td>
<td>iSTART</td>
<td>1.83 (.46)</td>
<td>2.06 (.48)</td>
<td>0.87</td>
</tr>
</tbody>
</table>

Notes: IP= Interactive Paraphrasing; * p = .08; ** p < .05

Discussion

We developed IP because prior experiments have shown that paraphrasing is a critical component of self-explanation, providing the foundation for the self-explanation, and because less-skilled readers tend to improve in paraphrasing and show less improvement in other strategies. IP was tested with high school students by examining before and after training, the students’ self-explanation quality, strategy use during comprehension, and their on-line ability to comprehend text. We found that iSTART without IP training was most effective for less-skilled readers (based on self-explanation quality, see also [14]). In contrast, IP-iSTART was more beneficial to skilled readers. The finding that IP-iSTART benefited skilled readers more so than less-skilled readers may appear somewhat surprising. Given that paraphrasing is the foundation for self-explanation, IP-iSTART should provide a scaffold for training less-skilled readers on how to use this strategy prior to training on the more challenging inference strategies associated with self-explanation. Further, skilled readers should already be adept at paraphrasing and shouldn’t need or benefit from this module. Nonetheless, these results show that, skilled (high school aged) readers can benefit from the additional scaffolding that the paraphrase practice provided. One distinct possibility is that the additional paraphrasing practice helped more skilled readers to better distinguish paraphrasing from self-explanation. This may be an emerging, but fuzzy, differentiation for skilled readers in the iSTART condition who did not receive the additional practice.

More difficult to answer is why the less-skilled readers did not improve from IP-iSTART training, whereas they did improve as a function of iSTART training without paraphrase practice. The latter finding is not surprising because we have found before that less skilled readers benefit from iSTART training. However, we found here that the additional focus on paraphrasing did not help the less skilled readers. One possibility is that the less-skilled students may have been overwhelmed and perhaps confused by receiving paraphrase practice during the introduction to the strategies, which is geared toward teaching them how to self-explain using a variety of strategies (not just paraphrasing). Then, during the self-explanation practice, they may have been discouraged from paraphrasing, which contradicts the paraphrase practice received during the introduction phase. Whereas the less-skilled readers may have been confused, the more skilled readers may have discerned the purpose and the difference between the two separate practice modules. For that reason, we are not yet convinced that less-skilled readers cannot benefit from paraphrase training. Hence, our current
efforts are geared toward improving the paraphrase module and discerning whether the paraphrase training should be provided in a separate module for less-skilled readers. Specifically, we are developing a separate module for less-skilled readers that will explain what reading to understand entails and we are integrating IP training into that module. Our expectation is that making improvements to the IP module and taking it out of the context of self-explanation will help the less-skilled readers to gain experience from paraphrase practice.

References


[17] T. O’Reilly & D.S. McNamara, Reversing the reverse cohesion effect: Good texts can be better for Strategic, high-knowledge readers, Discourse Processes 43 (2007), 121-152.

