January 2012

Changing How Students Process and Comprehend Texts with Computer-based Self-Explanation Training

Christopher A. Kurby
Grand Valley State University

Joseph P. Magliano
Northern Illinois University

Srikanth Dandotkar
Eastern Illinois University, sdandotkar@eiu.edu

James Woehrle
Northern Illinois University

Sara Gilliam
Northern Illinois University

See next page for additional authors

Authors
Christopher A. Kurby, Joseph P. Magliano, Srikanth Dandotkar, James Woehrle, Sara Gilliam, and Danielle S. McNamara

Changing How Students Process and Comprehend Texts with Computer-based Self-Explanation Training

Christopher A. Kurby

Grand Valley State University

Joseph P. Magliano, Srikanth Dandotkar, James Woehrle, Sara Gilliam,

Northern Illinois University

Danielle S. McNamara

The University of Memphis

Please address correspondence to:

Christopher A. Kurby

Grand Valley State University

Department of Psychology

2224 Au Sable Hall

One Campus Dr.

Allendale, MI 49401

kurbyc@gvsu.edu

Abstract

This study assessed whether and how self-explanation reading training, provided by iSTART (Interactive Strategy Training for Active Reading and Thinking), improves the effectiveness of comprehension processes. iSTART teaches students how to self-explain and which strategies will most effectively aid comprehension from moment-to-moment. We used RSAT (Reading Strategy Assessment Tool) to assess how iSTART changes the relation between important self-explanation reading strategies—bridging and elaboration—and online comprehension, and how often they are produced. College and high school students received iSTART and were administered RSAT prior to and post-training. Results from three experiments showed that iSTART primarily benefits bridging inferences when self explaining. The frequency of bridging inferences was higher post training than prior to training, but only in the experiments involving college students. Additionally, prior to exposure to iSTART, RSAT bridging scores did not predict comprehension performance, whereas they did after iSTART, suggesting that iSTART may improve comprehension processes by teaching students how to appropriately use self-explanation to address comprehension difficulties. Finally, the results from this study suggest that RSAT may provide a valuable computer-based assessment of the effectiveness of self-explanations that could be used in conjunction with iSTART and in future research on self-explanation.

Changing How Students Process and Comprehend Texts with Computer-based Self-Explanation Training

Success in secondary and post-secondary education is contingent on students' ability to understand and reason with the texts used in their courses. However, there is little to no classroom time spent on teaching students strategies that help them learn from and with challenging texts at this stage of education (Snow, 2002). One reason for this is that content instructors tend not to devote classroom time to teaching literacy practices that will help students better comprehend the challenging texts that face their students. One potential solution to this issue is to rely on intelligent tutoring systems that can provide individualized training to students outside the classroom (e.g., McNamara, 2009; McNamara, Levinstein, & Boonthum, 2004).

However, to provide automated strategy training, strategies must be identified that are both effective in promoting comprehension and amenable to computer-based training. One such strategy is self-explanation (McNamara et al., 2004), which is the practice of articulating and explaining to oneself the information derived from the discourse (Chi, Bassok, Lewis, Reimann & Glaser, 1989; Chi, De Leeuw, Chiu, & LaVancher, 1994; de Leeuw, & Chi, 2003; McNamara, 2004; McNamara & Magliano, 2009a; McNamara & Scott, 1999). When a studentstudents self-explains material, they draw upon information from the discourse context and relevant world knowledge in order to answer "why" questions posed to themselves (e.g., "why is this event occurring?" "why is the author mentioning this information?" "why am I not understanding?").

Successful self-explanation requires the metacognitive awareness to recognize when comprehension goals are not met or when one is failing to comprehend the text (Chi et al., 1989; de Leeuw, & Chi, 2003; McNamara & Magliano, 2009a; McNamara, O'Reilly, Rowe, Boonthum, & Levinstein, 2007). Moreover, self-explanation should vary given the demands of the texts (McNamara & Magliano, 2009a). For example, at one location in a text, an appropriate self-explanation may rely heavily on a causal process that is clearly delineated in the prior discourse contexts, whereas at another location, a reader may have to rely on text-relevant background knowledge. Accordingly, the goal of self-explanation training is two-fold: 1) to have students produce more self-explanations, and 2) to teach students when it is appropriate to employ self-explanation strategies and what kinds of self-explanations are most appropriate (Levinstein, Boonthum, Pillarisetti, Bell, & McNamara, 2007; McNamara et al., 2004).

Of particular interest in the current study is a computer-based reading strategy tutor, called iSTART (Interactive Strategy Trainer for Active Reading and Thinking; McNamara et al., 2004). iSTART decomposes self-explanation into a set of sub-strategies, which include paraphrasing what was just read, bridging to prior discourse, elaborating based on text-relevant background knowledge, and anticipating what will happen next. A central assumption of iSTART is that breaking self-explanation into sub-strategies will not only make the strategy more “concrete” to the student, but also help them develop an understanding of when each of these strategies will be most effective (McNamara, 2004). Students using iSTART interact with a set of animated agents that introduce the strategies, demonstrate their use, and guide the students as they practice self-explaining the texts. Students are taught two important skills associated with effective self-explanation. The first involves the ability to recognize how to discriminate between the strategies and the second is to recognize when it is appropriate to use them. For example, if a student recognizes that she does not understand why an event is happening and there is an answer in the prior discourse context, then she needs to locate that information to effectively explain the sentence. If the text invites the reader to reason beyond the text, she must recognize that this is an appropriate point at which to base her self-explanation on relevant world knowledge.

What evidence is there that self-explanation improves comprehension and can be taught? First, readers who spontaneously self-explain the text typically comprehend better than those who do not (Chi et al., 1989; Trabasso & Magliano, 1996; Coté, & Goldman, 1999; Coté, & Goldman, & Saul, 1998). Second, better quality of self-explanations correlate with better scores on comprehension tests (Ozuru, Briner, Best, & McNamara, 2010). Finally, instructing self-explanation strategies through classroom interventions or through iSTART improves the quality of self-explanations, and improves performance on offline comprehension tests, compared to control conditions in which participants did not receive training (Magliano, Todaro, Millis, Wiemer-Hastings, Kim, & McNamara, 2005; McNamara, 2004; McNamara, O'Reilly, Best, & Ozuru, 2006; O'Reilly, Best, & McNamara, 2004; O'Reilly, Taylor, & McNamara, 2006). iSTART, the computer-based version of training, has been shown to both increase the frequency of strategies that comprise self-explanations and improve comprehension (Magliano et al., 2005; McNamara et al., 2004; McNamara et al., 2006). In addition to increasing the frequency of self-explanation, iSTART has been argued to increase the alignment between self-explanation processes and comprehension. After training, the quality and use of self-explanation strategies should be more systematically related to comprehension performance, compared to before training (McNamara & Magliano, 2009a).

Although self-explanation and reading strategy interventions can improve text understanding, the mechanisms underlying this change have not been thoroughly explored. If training improves the effective use of strategies, we should see better alignment between the reading processes in which readers engage and comprehension. That is, we should find evidence that strategies promoted by iSTART support comprehension more so after training than before training. iSTART emphasizes two strategies that are particularly important in self-explanation: bridging and elaboration. Research has shown that bridging inferences that link back to the prior text are necessary to maintain globally coherent representations of discourse, and that they are a normal concomitant of successful comprehension (see Graesser, Singer, & Trabasso, 1994, for an extensive review). Elaborations allow the reader to embellish upon their understanding, and make links between textual information and knowledge structures stored in long-term memory (e.g., Kintsch, 1998; Spilich, Vesonder, Chiesi, & Voss, 1979). These inferences are a characteristic of high ability readers and contribute to readers' ability to understand text (Long, Oppy, & Seely, 1994; McNamara & McDaniel, 2004; Oakhill & Yuill, 1996). **Typically, the more bridging inferences and elaborations present in a self-explanation, the better the quality (Magliano, Millis, The RSAT Development Team, Levinstein, & Boonthum, 2011ref).

The general goal of this study is to better understand how self-explanations and self-explanation training contribute to comprehension. To this end, the current study was conducted to answer the questions of (1) whether iSTART training increases the use of reading strategies important to self-explanation, and (2) whether iSTART better aligns self-explanation processes with comprehension after training. This second question is motivated by the general assumption that iSTART works because it helps foster the metacognitive awareness of when the sub-strategies emphasized within the training are more critical for comprehension (McNamara & Magliano, 2009a; McNamara et al., 2007). It is important to note that the answers to the above two research questions are not likely to be mutually exclusive and are not treated as such in this study.

To assess the change in self-explanation usage and the alignment between self-explanation processes and comprehension, we used an automated reading strategy assessment tool called the Reading Strategy Assessment Tool (RSAT; Gilliam, Magliano, Millis, Levinstein, & Boonthum, 2007; Magliano, et al., 2011). RSAT was developed to provide a computer-based approach for collecting and analyzing verbal protocols produced while reading texts. It is intended to provide assessments of comprehension as well as the processes that give rise to it. RSAT requires the student to read texts on a computer and answer open-ended questions that are embedded within the texts. After reading pre-selected target sentences, readers are asked to produce responses to indirect and direct questions. Indirect questions are intended to tap comprehension processes and require readers to report thoughts regarding their understanding of the sentence in the context of the passage (Instructions include "What are you thinking now?"). More specifically, these responses are analyzed for the presence of paraphrases, bridges, and elaborations, which are the strategies taught in iSTART. Direct questions are designed to assess comprehension level and require readers to answer specific "wh-" questions about the text at target sentences (e.g., "Why was the Union demoralized?" in a passage about the American Civil War). Direct questions are designed to provide a direct assessment of *emerging comprehension* – that is, comprehension as students are reading a text. Performance on direct questions could provide a basis for assessing comprehension skill that is an alternative to standardized comprehension tests that base assessment on multiple choice tests that are subject to specialized test taking strategies (e.g., Magliano et al., in press2011)

RSAT uses automated scoring procedures, resulting in sets of word counts, to measure comprehension and use of reading strategies. RSAT has been shown to have respectable validity (Magliano et al., in press2011) and reliability (Millis & Magliano, in press). First, RSAT comprehension (based on word counts for direct questions), ACT, and Gates-MacGinitie scores are all comparably correlated with one another. Second, RSAT process measures show convergent validity in that for paraphrasing, bridging, and elaboration processing scores are correlated (Pearson r ranging from .48-.75) with human judgments of those strategies. Correlations between test-retest scores for different forms of RSAT are also respectable given the open ended nature of the task of "thinking aloud" (Pearson r ranging from .55-.79). It is important to note that text characteristics affect the processes in which readers engage while reading, which vary from text to text (e.g., Magliano, Trabasso, & Graesser, 1999). Additionally, we have observed that students tend to decrease the amount of verbiage across administrations of a long form of RSAT (Millis & Magliano, in press). Millis & Magliano (in press) speculate that this may be due to participant fatigue or a practice effect, however, more research needs to be done regarding this issue. Regardless, complicates use of a long form in a study exploring changes in strategy use as a function of strategy training. This later finding motivated the development of a short form of RSAT (Magliano, Millis, & LevensteinLevinstein, 2010), which is assessed in the current study (Experiment 2).

A central thesis of the current study is that iSTART has two effects on the comprehension process: 1) it increases the frequency of the use of self-explanation strategies, and 2) it helps students learn when to use these strategies, resulting in a better alignment between self-explanation strategies and comprehension. We used various versions of RSAT in two different populations (college and high school), across three experiments, to address these issues. In Experiment 1, we used a long form of RSAT, which contained six texts, to assess strategy use after iSTART training compared to a control condition. As mentioned earlier, there is a reduction in verbiage with multiple exposures to RSAT, likely due to fatigue induced from the procedure

(Millis & Magliano, in press). We observed such effects in the current study as well. Also, the long form of RSAT intermixes items designed to measure self-explanation strategies and comprehension performance. It is desirable to separate these two measures to gain independent assessments of them. Having separate assessments of processing and comprehension was necessary to evaluate the extent to which RSAT comprehension scores were correlated with individual differences in response to iSTART. In Experiment 2, we used short-form versions of RSAT to address both of these concerns. The short forms only contain two texts and there are separate forms to assess comprehension skill (RCAT) and the extent to which readers engage in comprehension processes (RSAT-S). In Experiment 3, we assessed the effectiveness of iSTART regarding self-explanation processes and alignment in a high school sample using a shortened, and modified, version of RSAT that contained both comprehension and processing items.

We make two predictions regarding the effectiveness of iSTART. First, if iSTART increases the use of self-explanation strategies after training, then the frequency of bridging and elaboration, as measured with RSAT, should be higher for those trained than those untrained. Second, if iSTART enables students to more effectively and appropriately engage in self-explanation, then the frequency of bridging and elaboration should be more strongly correlated with comprehension post-training than prior to training.

Experiment 1

In this first experiment, participants from an undergraduate population were randomly assigned to an iSTART training condition and a control condition. In both conditions, participants were tested with a long form of RSAT twice, once before training, or a control activity, and once after training. RSAT provided a measure of paraphrasing, bridging, elaboration, and comprehension.

As mentioned earlier, the long form of RSAT has been found to reduce the amount of verbiage after repeated testing (Magliano Millis & Magliano, in press), which we also observed in the current study. This decrease would compromise the ability to detect changes in strategies comparing the pre- and post-training sessions of RSAT. Hence, to assess the effects of iSTART on the frequency of self-explanation, Analysis of Covariance (ANCOVA) was used, controlling for use of the strategies prior to training. Specifically, we conducted separate ANCOVAs on the frequency of paraphrases, bridges, and elaborations post-training comparing iSTART and control, controlling for the frequency of producing each strategy pre-training. These analyses reveal differences in strategy use between the iSTART and control conditions, over and above what students do prior to training. Previous research has shown that individual differences in comprehension ability predict responsiveness to self-explanation training (Magliano, Wiemer-Hastings, Millis, Munoz, & McNamara, 2002). As such, we also controlled for pre-existing differences in comprehension ability using the American College Test (ACT).

Finally, we assessed the alignment between self-explanation strategies and comprehension by conducting multiple regression analyses using the RSAT processing measures of paraphrasing, bridging, and elaboration to predict comprehension measured by the RSAT comprehension measure (i.e., direct questions). RSAT processing measures differentially correlate with the comprehension scores (Magliano et al., in press2011). The correlations are indicative of the extent to which those processes are supporting comprehension for a given reader. Therefore, we hypothesized that changes in the correlations from pre to post training would be indicative of changes in how readers are supporting comprehension after training (see also, McNamara et al., 2006).

Methods

Participants. One hundred and forty college students enrolled in a critical thinking course at Northern Illinois University participated for course credit and in particular, two sections of this course taught by the same instructor. The sections were randomly assigned to the iSTART (66 students) or control condition (74 students). It is well documented that response to strategy training is moderated by individual difference factors, such as comprehension skills (e.g., Magliano et al., 2005). Therefore, we asked permission to gain access to ACT scores. Forty-seven students in the iSTART condition and 44 students in the control conditions had ACT scores that were accessible. All analyses used only these participants. Mean ACT performance on the comprehension test was 22.84 (*SD* = 3.39), which was similar to the national norm for 2009 (*M* = 21.1; ACT National Profile Report, 2009).

RSAT. RSAT (Gilliam et al. 2007; Magliano et al., in press2011) was administered on personal computers in a web-based environment. The texts are presented in black font in a gray field left justified near the top of the computer screen. The title of each text remained centered at the top of the screen while participants read the entire text. In the current study, only one sentence of a text was shown on the screen during reading. Participants navigated forward through the text by clicking on a “next” button, which is located near the bottom left portion of the computer screen. Participants could not move backwards through the text at any point. “NEW PARAGRAPH” markers appeared when there is a shift to a new paragraph. After participants clicked the “next” button, the next sentence appeared, provided it was a non-target sentence. For target sentences, a response box appeared to the right of the “next” button with a prompt above the box. The prompt for an indirect question was “*What are you thinking now*?” For direct questions, the target sentence was removed from the screen when the question and response box appeared. Participants typed their answers to the question in the response box.

They clicked the next button when they were finished, after which the response box disappeared and the next sentence was presented. The order of the texts was randomly presented to the participants.

Each answer to the target sentences was automatically scored by identifying the number of content words in the answer that was also in the text or in an ideal answer (Gilliam et al., 2007; Magliano et al., in press2011). Content words included nouns, adverbs, adjectives and verbs (semantically depleted verbs, such as *is, are*, were omitted). Word matching was accomplished by literal word matching and Soundex matching, which detected misspellings and changes in verb forms (Birtwisle, 2002; Christian, 1998). For answers to the indirect question, four scores were computed. The *paraphrase* score was the number of content words from the target sentence. RSAT computes both *local* and *distal-bridging scores*. Local-bridging scores reflect the extent to which the verbal protocols contain content words form the immediate prior sentence and *distal-bridging* scores comprise the number of content words from all other sentences that have been read thus far. The *elaboration* score was the number of content words in the answer that were not present in the prior discourse context. For the direct questions, there was only one score computed: the number of content words in the answer that was in the ideal answer. For each participant, we computed mean scores by averaging over the individual scores obtained for each target sentence. Therefore, we calculated mean scores for paraphrases, local and distal bridges, elaborations from the answers to the indirect questions, and mean comprehension scores from the answers to the direct question. Tables 1 and 2 present example protocols and their scoring. Also included in the tables are the word matches from each protocol that determined the different RSAT scores. Table 1 presents examples for an indirect question, and Table 2 presents examples for a direct question.

-- INSERT TABLE 1 HERE --

-- INSERT TABLE 2 HERE --

Two stimulus lists of passages were used in RSAT based on prior research (Magliano et al., in press). Each stimulus set contained six texts: two science texts, two history texts, and two narratives. The sentences where the indirect and direct questions occurred were determined based on prior research (Gilliam et al., 2007; Magliano et al., in press2011). The presentation of the stimulus list was counterbalanced across the first and second administrations of RSAT.

iSTART. Participants received the version of iSTART described in Levinstein et al., 2007). Briefly, iSTART teaches self-explanation and reading strategies and consists of three sections: *introduction, demonstration,* and *practice*. The introduction uses three animated agents to simulate a classroom type of discussion between a teacher and two students. They define and provide examples of self-explanation and the reading strategies, including comprehension monitoring, paraphrasing, predicting, making bridging inferences, and elaboration. The demonstration module shows the use of the strategies through the interaction of teacher and student computer agents. The practice section allows students to practice the newly learned strategies during reading. The self-explanations that students generate in the practice module are automatically evaluated and immediate feedback is provided to the participant (see McNamara et al., 2007, for more details).

Procedure. RSAT and iSTART sessions were administered in the context of course laboratory exercises. The first session of RSAT occurred during the first two weeks of classes. The procedures for administering RSAT developed by Magliano et al. (in press) were adopted. These procedures consist of giving general instructions regarding how to respond to indirect and direct prompts. With respect to the indirect prompts, participants were told to produce thoughts that corresponded to how they understood the sentence they just read in the context of the passages (see also Trabasso & Magliano, 1996). With respect to the direct prompts, participants were told to answer the questions as completely as possible. Participants were given a pen and paper practice test that consisted of a short five sentence text containing two indirect prompts and on direct prompts. The experimenters reviewed the participants' responses for the practice passage and gave general feedback. If participants produced vague and uninformative responses to the indirect prompts, the participant reiterated the instructions for these prompts. Participants in the iSTART condition were administered iSTART training during the third and fourth weeks of the class, which overlapped with lectures on reading comprehension. They were administered iSTART over two sessions, which lasted approximately an hour each. The first session consisted of the introduction and demonstration phases of iSTART and the second session consisted of the practice phase. The control condition simply attended lecture during those days. During the fifth and sixth weeks of the course, both conditions were administered RSAT a second time.

Results and Discussion

There were three sets of analyses. The first was conducted to ensure that there were no pre-existing differences between the iSTART and control conditions with respect to the pre-training RSAT processing and comprehension scores. The second set was conducted on the post training scores and consisted of a series of ANCOVAs in which the pre-training RSAT scores and performance on the ACT served as covariates. The third set involved conducting multiple regression analyses to assess the extent to which the RSAT processing scores were correlated with the RSAT comprehension scores prior to and post training.

Effects of iSTART on the frequency of strategy production and comprehension. Table presents the means and standard errors for the RSAT scores as a function of training condition and session (pre- vs. post-training). For the first set of analyses, we conducted a separate ANCOVA, with training condition as the independent variable (iSTART vs. control) and ACT as the covariate, on each pre-training RSAT processing score: 1) paraphrasing, 2) bridging (the analysis on bridging scores used local and distal bridging as a within-participants factor, referred to as *distance*), 3) elaboration, and 4) the RSAT comprehension scores. None of the analyses revealed significant differences across the training conditions (all $p > 0.10$), indicating that there were no differences in processing and comprehension between the iSTART and control conditions prior to exposure to iSTART.

-- INSERT TABLE 3 HERE --

The next set of ANCOVAs was conducted on post-training scores. As was the case with pre-training scores, separate analyses were conducted on each strategy and the comprehension score using training condition as the independent variable and ACT score as a covariate. In addition, the corresponding pre-training score was added as a covariate to control for the use of the strategy prior to training (e.g., the pre-training paraphrase score was used as a covariate for the analysis on the post-training paraphrase scores). Table 3 contains the mean adjusted post-training strategy scores as a function of condition.

The analyses conducted on the post-training paraphrase and elaboration scores did not reveal a significant difference as a function of strategy training (both $F < 1$). By contrast, the analysis on bridging scores revealed a marginally significant training X distance interaction, $F(1, 87) = 3.801$, $p = 0.054$, partial $\eta^2 = .046$. Follow-up t-tests revealed that there were no differences between the training conditions with respect to local bridging scores ($p = .46$), but that students receiving iSTART had higher distal bridging scores than those not receiving iSTART ($d = .49$, $p = .02$). Finally, the analysis on the RSAT comprehension score did not reveal a difference between training conditions ($p > 0.10$). These analyses are more or less consistent with prior research that shows that self-explanation training primarily increases the use of bridging.

Effects of iSTART on alignment between strategies and comprehension. The final set of analyses in Experiment 1 consisted of regression analyses using the RSAT processing measures to predict RSAT comprehension. Specifically, mean paraphrasing, bridging, and, elaboration scores were used to predict performance on comprehension as measured by performance on the direct questions (i.e., RSAT comprehension score). Additionally, the ACT scores were entered into the equation to control for general comprehension skill. We computed a total bridging score by adding local and distal bridging scores. Separate analyses were conducted on the pre- and post-training sessions and for the control and iSTART conditions. These analyses allowed an assessment of whether exposure to training changes how processing supports comprehension. All predictors were simultaneously force-entered into the regression equations.

First, consider the regression analyses for the students receiving iSTART. For the pre-training scores, the RSAT processing variables accounted for a significant 46% of the variance in the RSAT comprehension scores, $F(4, 42) = 8.95$, $p < .001$. As shown in Table 4, of the RSAT measures, paraphrase and elaboration scores significantly predicted comprehension scores. It is important to note that we have found in other samples that distal bridging and elaboration scores are significant and positive predictors of RSAT comprehension scores without iSTART training (Magliano et al., 2009). Indeed, one advantage of using RSAT in this context is that it provides baseline information regarding how a particular sample of students (or individual) comprehends what they read prior to training.

-- INSERT TABLE 4 HERE --

The regression analysis for the post-training scores accounted for a significant 61% of the variance, $F(4, 42) = 16.68$, $p < .001$ (see Table 4). After training, bridging scores were significant and positive predictors of comprehension, and elaboration scores approached significance and again were also positively correlated with comprehension. The primary difference between pre and post-training analyses was that the bridging scores were significant in the latter, but not in the former analyses. These analyses converge with the ANCOVA on post strategy bridging scores, which suggest that iSTART training promotes bridging and its effective use in supporting comprehension.

Next, consider the regression analyses for the students in the control condition. For the pre-training scores, the RSAT measures accounted for a significant 44% of the variance of the RSAT comprehension score, $F(4, 39) = 7.50$, $p < .001$. As shown in Table 8, of the RSAT measures, bridging scores significantly predicted comprehension scores. The regression analysis for the post-training scores accounted for a significant 39% of the variance, $F(4, 39) = 5.62$, $p < .001$], which is comparable to the analyses on the pre-training protocols. In the second administration of RSAT, both paraphrasing and bridging scores were approaching significance and were positive predictors of comprehension. These data suggest that the pattern of significance for the control condition was more or less consistent across analyses, with the exception that paraphrasing scores approached significance in the analysis on the second administration of RSAT for these participants.

Summary

These results suggest that iSTART both changes the frequency of self-explanation strategies, in this case distal bridging, and also increased the alignment between self-explanation processes and comprehension. This supports the claim that iSTART improves both how often readers engage in self-explanation and how well self-explanation is used (McNamara & Magliano, 2009a). Additionally, these results show the utility of an automated coding system, such as RSAT, in conducting research of this nature. The challenges of hand coding severely limits sample sizes that can be readily used in a study of this nature. However, the general problem of the decrease in verbiage found in the long form, and as is evident in Table 3, is a problem for the utility of this tool. Therefore, Experiment 2 used a short form that was developed by Magliano et al. (2010).

Experiment 2

Experiment 2 was conducted as a replication of Experiment 1, with some modifications. The modifications addressed some challenges posed by the long form of RSAT, used in Experiment 1, in the assessment of changes in processing due to iSTART. The first of the challenges is that verbiage generally decreased from pre-training to post-training, which makes it impossible to assess changes in the production of reading strategies from pre-training to post-training. In Experiment 2, we took advantage of a short form of RSAT, which contains only two of the texts from the long form. This approach was expected to reduce fatigue effects that may occur with the long form of RSAT. A second challenge posed by the RSAT long form is that the indirect and direct questions were intermixed within each text and so one would expect processing and comprehension measures to be correlated. A stronger test would be to have processing and comprehension measures coming from different texts.

Additionally, we intended to use RSAT comprehension scores to explore individual differences in response to iSTART exposure because we have shown that readers of varying skill respond differently to self-explanation training (e.g., Magliano et al., 2005; McNamara et al., 2006). For example, Magliano et al. (2005) found that high ability readers tend to increase the quality of their self-explanations after iSTART training, although all readers increased their reliance on text-based information when producing self-explanations. McNamara et al. (2006) found that readers with low prior knowledge about reading strategies showed improved text-based processing after iSTART training whereas high strategy knowledge readers showed improved coherence-building processes, as revealed by bridging performance. These studies show that it is ideal to model individual differences when investigating the effectiveness of iSTART on reading comprehension. Give that one motivation of this study is to assess the utility of RSAT, we wanted to use the comprehension measure associated with it. It is not tenable, however, to use the comprehension score based on the RSAT long form to explore individual differences because the processing and comprehension measure share variance. It is desirable to reduce this shared method variance and obtain independent measures of these two types of processes.

Magliano et al. (2010) developed a short form of RSAT that contains only direct questions and a short form of RSAT that contains only indirect questions. The modified version of RSAT, called the Reading Comprehension Assessment Tool (RCAT), which only contained direct questions allowed us to obtain an independent measure of quality and individual differences in comprehension. RSAT-S, a short form of RSAT that only contains indirect questions, was used to assess processing measures. We had two forms of each measure.

Methods

Participants. Similar recruitment strategies were used as in Experiment 1. One hundred forty students participated in this experiment. All students were enrolled in a Critical Thinking course at Northern Illinois University. One section of the course, consisting of 72 students, was randomly assigned to receive iSTART, and one section, consisting of 68 students, was randomly assigned to the control condition.

RSAT and RCAT. RSAT was administered using the same methods as described in Experiment 1, with the exception that the short form was used. The short forms (RSAT-S and RCAT) consisted of two texts selected by determining which texts in the long form contained items most correlated with independent measures of comprehension (Magliano et al., 2010). RSAT-S included only indirect questions. RSAT scores of reading processes (e.g., paraphrasing, bridging, elaboration) were computed using the same procedures as in Experiment 1. The presentation of RCAT was the same as RSAT with the exception that all questions were direct questions. Performance on RCAT was computed in the same way as was performance on direct questions in Experiment 1: the average number of words that overlapped between each given answer and ideal answer.

Two stimulus lists were created for RSAT and RCAT, creating two forms of each. Each form of RSAT and RCAT contained two texts, a science text and a history text. The assignment of form to pre-training and post-training was counterbalanced across participants.

iSTART. The same version of iSTART was used in this experiment as in Experiment 1.

Procedure. RSAT, RCAT, and iSTART were administered in the context of course laboratory exercises, as in Experiment 1. Participants were administered one of two forms of RSAT and RCAT during the first two weeks of the course, which did not focus on reading comprehension. Participants in the iSTART condition were administered iSTART training during the third and fourth weeks of the class, which overlapped with lectures on reading comprehension. They were administered iSTART over two sessions, which lasted approximately an hour each. The control class only attended lecture during this time. During the fifth and sixth weeks of the course, both sections were administered RSAT and RCAT a second time. The two forms of RSAT and RCAT were counterbalanced across the two sessions.

Results and Discussion

There were two sets of analyses. The first involved assessing changes in strategies and comprehension performance as a function of session, training condition, and individual differences in comprehension skill, and the second assessed the alignment between comprehension processes strategies and comprehension performance.

Effects of iSTART on the frequency of strategy production and comprehension. The first set of analyses examined the effects of iSTART on the frequency of use of the different reading processes measured by RSAT-S. We also assessed whether these effects were moderated by individual differences in comprehension skill, as measured by RCAT. As such, we divided participants into skilled and less-skilled groups by using a median split on overall RCAT scores. Table 5 provides the means and standard errors for each condition.

-- INSERT TABLE 5 HERE --

As was the case in Experiment 1, the first analysis was conducted on the pre-training RSAT scores to determine that there were no differences between participants in the training conditions prior to exposure to iSTART. Additionally, the inclusion of the RSAT comprehension score enabled us to assess the validity of this measure because scores should be different between skilled and less skilled readers. A series of training (iSTART vs. control) X comprehension skill (low vs. high) ANOVAs were conducted on 1) paraphrase scores, 2) bridging scores, and 3) elaboration scores. The analysis on paraphrase scores revealed a main effect of comprehension skill, such that high skilled readers had higher scores ($M = 1.79$, $SE = .09$) than low skilled readers ($M = 1.26$, $SE = .09$), $F(1, 131) = 19.59$, $MSE = 0.48$, $p < .001$, partial η^2 = .127. Similarly, the analysis on bridging scores score revealed a main effect of comprehension skill, such that high skilled readers had higher scores (*M* = 1.95, *SE* = .17) than low skilled readers (*M* = 2.97, *SE* = .16), *F*(1, 131) = 19.76, *MSE* = 1.81, $p < .001$, partial η^2 = .128, Finally, again, the analysis on elaboration scores revealed a main effect of comprehension skill, such that high skilled readers had higher scores (*M* = 5.89, *SE* = .34) than low skilled readers (*M* = 4.43, *SE* = .36), *F*(1, 131) = 8.43, *MSE* = 8.68, *p* = .004, partial η^2 = .059. Most importantly, there were no effects involving training condition, indicating that both reading skill groups were statistically equivalent prior to training (all $p > .20$)

Next, for each RSAT measure, we conducted a mixed ANOVA with training condition (iSTART vs. control) and comprehension skill (skilled vs. less skilled) as between participant variables, and session (pre- vs. post-training) as a within-participants variable. Given that we were primarily interested in effects that involve training condition, we report only those. For paraphrasing, the there were no significant effects involving training condition. There were no other significant effects (all $p > .10$).

For bridging, in contrast to Experiment 1, we combined local and distal bridging scores to reduce the number of factors in the ANOVA. The ANOVA revealed a significant interaction among training condition, session, and comprehension skill, *F*(1, 135) = 6.94, *MSE* = 0.68, *p* = .009, partial η^2 = .046. Two follow-up ANOVAs were conducted for each comprehension skill level. For less skilled readers, a training X session ANOVA revealed no significant effects (all p > .20). However, the same ANOVA for skilled readers revealed a significant training X session interaction, *F*(1,72) = 6.92, *MSE* = 0.88, *p* = .01, partial η^2 = .088. We conducted post hoc t-tests comparing differences between sessions as a function of training. As is evident from the means and standard errors in Table 5, only the skilled readers receiving iSTART had higher bridging scores in the post-training than the pre-training scores, $t(36) = 3.83$, $d = .63$, $p < 0.001$. This reading skill by training interaction replicates Magliano et al. (2005) and McNamara et al. (2006) but with an automated assessment of bridging. There were no other significant effects (all p's > .10).

For elaboration, the ANOVA revealed a significant interaction among training, session, and comprehension skill, $F(1, 135) = 5.08$, $MSE = 4.95$, $\eta^2 = .035$, $p = .028$. Two follow-up ANOVAs were conducted for each comprehension skill level. There were no significant effects involving training condition for less skilled readers (all $p > .20$). The ANOVA for skilled readers revealed a significant training X session interaction, $F(1, 72) = 5.80$, $MSE = 3.24$, $\eta^2 = .075$, $p = .019$. Follow up tests revealed that skilled readers in the control condition had significantly lower elaboration scores in the second session of RSAT than the first, $t(37) = 4.77$, $d = .72$, $p > .001$, but the difference between pre-training and post-training scores for the iSTART condition was not significant.

Summary

First, these results indicate that the short forms of RSAT are more appropriate and effective in evaluating changes in self-explanation processes as a function of exposure to iSTART than the long forms. The clearest and interpretable results pertain to paraphrasing and bridging. Students receiving iSTART paraphrased more after training than before, whereas there were no changes for the students not receiving iSTART.

Second, with respect to bridging, only the skilled readers showed an increase in bridging as a function of iSTART. These findings of individual differences in responsiveness to iSTART are consistent with previous work (Magliano et al., 2005; McNamara et al., 2006). The results suggest that all readers became more focused on the text content after training, and that high ability readers increased the integrative processing they perform on this content. Less-skilled readers typically produce ineffectual elaborations before self-explanation training, and less so following training (Bellissens, Jeuniaux, Duran, McNamara, 2009). As such, it is desirable to improve processing of the text in low-ability readers in order to buttress more complex comprehension operations. Indeed our data show that iSTART served to increase paraphrasing but reduce elaborations in low-ability readers. Future work is aimed at improving bridging performance in low ability readers (Jackson, Demspey & McNamara, in press). It is encouraging that these effects can be detected using RSAT, which suggests that automated coding of self-explanation protocols is a viable way for continuing this line of research.

Effects of iSTART on alignment between strategies and comprehension. Similar to Experiment 1, we conducted a series of regression analyses to assess the extent to which RSAT processing measures accounted for variance in comprehension performance (as measured by RCAT) prior to and post training. (It is important to note that we did not have ACT scores for this experiment as we did for Experiment 1.) We conducted four multiple regression analyses, one on the pre-training scores, and one on the post-training scores, for students in the iSTART and control conditions. The RCAT comprehension score was the criterion variable, and the paraphrase, bridging, and elaboration scores were the predictor variables. The predictors were simultaneously force-entered into the regression equation.

Table 6 presents the results of these analyses. First, consider the regression analyses for the students receiving iSTART. For the pre-training scores, RSAT variables accounted for a significant 43% of the variance in comprehension as measured by RCAT, $F(3, 68) = 17.31$, $p < .001$. Of the RSAT measures, paraphrase and elaboration scores significantly predicted comprehension scores; greater production of paraphrases and elaborations was associated with better comprehension. The regression analysis for the post-training scores accounted for a significant 40% of the variance, $F(3, 68) = 15.08$, $p < .001$. After training, only bridging scores were significant and positive predictors of comprehension. The primary difference between pre and post-training analyses was that the bridging scores were significant in the latter, but not the former analyses.

-- INSERT TABLE 6 HERE --

Next, consider the regression analyses for the students in the control condition. For the pre-training scores, RSAT variables accounted for a significant 20% of the variance in comprehension, $F(3, 64) = 8.67$, $p < .001$. As shown in Table 6, of the RSAT measures, bridging scores significantly predicted comprehension scores. The regression analysis for the post-training scores accounted for a significant 29% of the variance, $F(3, 64) = 5.39$, $p < .001$. In the second administration of RSAT, as consistent with the pre-training analyses, only bridging scores were significantly correlated with comprehension performance. Thus, the alignment between self-explanation processes and comprehension did not change across the two testing sessions.

These data are consistent with that involving the long form in Experiment 1, and indicate that exposure to iSTART increases the correlations between bridging and comprehension. These results suggest that iSTART increases the metacognitive awareness of the appropriateness of the different self-explanation strategies, and that RSAT is sensitive to such changes.

Experiment 3

The purpose of Experiment 3 was to assess whether these findings generalize to a high school population. However, it is important to note that there were some practical constraints that affected the design of the study. First, we were unable to recruit enough students to afford control and iSTART conditions. Second, there was a limited amount of time for each session, and as such we had to reduce the amount of time to administer RSAT and made additional modifications to the short form so that they contained both indirect and direct items. Although it is desirable to separate short forms as was done in Experiment 2, as will be evident, this shortened form of RSAT was successful in demonstrating some changes in text processing as a result of exposure to iSTART.

Methods

Participants. Thirty-five students from a Mid-South high school participated for $80 pay. Twenty-three students were age 16, 11 were 17, and one was 18. One student was a freshman, 30 were sophomores, and 4 were juniors. Twenty-eight students were female, and 7 were male.

Gates-MacGinitie Reading Test (GMRT). Reading skill was measured by the GMRT (3rd ed.) reading skill test (form L) level 7/9 (MacGinitie & MacGinitie, 1989). The GMRT consists of 48 multiple-choice questions designed to assess student comprehension on 14 short text passages. The test was time-limited to 20 minutes. Performance was calculated as the proportion of items answered correctly. Mean performance on the GMRT was .48 ($SD = 0.18$)

RSAT. There was a limitation in the length of time that we were able to have students participate in this study. As such, we had to reduce the time it required to administer RSAT. Rather than using RSAT-S and RCAT, as was the case in Experiment 2, we used the same text used in RSAT-S, but included both direct and indirect items. As such, this new form of RSAT provided both comprehension and processing scores. It is important to note that because this form of RSAT yielded both measures, we could not use the RSAT comprehension score as grouping or control variables for comprehension skill. This was the primary motivation for using performance on the Gates-MacGinitie test of comprehension.

Materials. Two stimulus lists were used in Study 2 for RSAT that consisted of four texts used in Study 1. Each stimulus list contained one science text and one history text and the texts in the lists were matched on length, difficulty, and number of indirect and direct prompts. The stimulus lists were counterbalanced across pre- and post-training sessions.

iSTART. Two different versions of iSTART were used for this experiment. One version was the same as used in Experiments 1 and 2. The other was a slightly modified version of iSTART to include extra instruction and practice in paraphrasing, which was embedded in the Introduction portion of iSTART (McNamara, Boonthum, Kurby, Magliano, Pillarisetti, & Bellissens, 2009). Half of the participants received one version, and the other half received the other. McNamara et al. (2009) found that this altered version of iSTART had only a small effect on self-explanation depending on reader ability. For the purposes of the present study, we dummy-coded iSTART version in our analyses to control for this variability.

Procedure. The experiment was divided into four short sessions and occurred over the course of two days. In the first session, which occurred on the first day, participants were administered the GMRT and RSAT. The GMRT was time-limited to 20 minutes, and RSAT was self-paced. The second, third, and fourth sessions occurred on the second day. In the second session, participants proceeded through the Introduction and Demonstration sections of iSTART. In the third session, participants engaged in the Practice module of iSTART. In the fourth session, participants received the other form of RSAT as a post-test.

Results and Discussion

There were two sets of analyses. The first involved assessing changes in strategies and comprehension performance as a function of training and the second assessed the alignment between comprehension processes and comprehension.

Effects of iSTART on the frequency of strategy production and comprehension. Table 7 contains the means and standard errors for the pre- and post-training session. Consistent with Experiments 1 and 2, we first assessed if there were pre-training differences as a function of training condition and conducted a series of ANCOVAS on the pre-training RSAT scores using performance on the GRMT as a covariate. There were no significant differences between the training conditions for any of the RSAT measures (all *ps* > .20).

-- INSERT TABLE 7 HERE --

As can be seen in Table 7, there were small changes in RSAT scores post-training. The paraphrase, bridging, and comprehension scores increased from pre-training to post-training, whereas the elaboration scores were essentially unchanged. We conducted a series of ANCOVAs on each RSAT score with session as a within-participants variable and performance on the GMRT as a covariate. None of the effects involving session were significant (all p > .10). However, it should be noted that there was a relatively low number of participants and as such, restricted power.

Effects of iSTART on alignment between strategies and comprehension. As was the case with in Experiments 1 and 2, we conducted two multiple regression analyses, one on the pre-training scores, and one on the post-training scores. Given that the participants did not have ACT scores, we used performance on the GMRT as our standardized comprehension score. The RSAT comprehension score was the criterion variable, and the paraphrase, local bridging, distal bridging, and elaboration and GMRT scores were the predictor variables. Additionally, the dummy coded variables were included to reflect the different forms of iSTART. The predictors were simultaneously force-entered into the regression equation.

Table 8 presents the results of these analyses. The pre-training regression accounted for 40% of the variance, $F(4, 44) = 9.62$, $SE = 0.61$, $p < .01$. As shown in Table 8, of the RSAT measures, only the paraphrase score significantly predicted comprehension scores after training. This result replicates the pre-training paraphrasing effect reported in Experiment 1.

-- INSERT TABLE 8 HERE --

The post-training regression accounted for 47% of the variance in comprehension scores, $F(5, 30) = 5.13$, $SE = 0.69$, $p < .01$, which, similar to Study 1, reflects an increase in the variance explained post-training relative to pre-training. As can be seen in Table 8, after training, of the RSAT measures, only the bridging score predicted comprehension scores. Paraphrase scores were no longer a significant predictor of comprehension scores. These results dovetail with those from Experiments 1 and 2 and show that although bridging was not related to comprehension prior to training, it significantly predicted comprehension after training.

The results of the regression analyses converge with those for Experiments 1 and 2. They paint a consistent picture that exposure to iSTART increases the extent to which bridging supports comprehension. Experiment 3 is notable because it replicates this finding in a population of high school students.

General Discussion

The goal of iSTART is to provide students with training on how to better self-explain texts (McNamara et al., 2004; McNamara et al., 2006). It accomplishes this objective by teaching students the sub-strategies of self-explanation. In this study, we conducted a direct test of whether iSTART training improves the alignment between comprehension and self-explanation using bridging and elaboration reading strategies. The results of all three experiments consistently showed that iSTART training primarily promotes bridging, of which the vast majority of discourse processing theories assume to be crucial for comprehension (e.g., McNamara & Magliano, 2009b). Experiments 1 and 2 showed increases in the frequency of bridging after exposure to iSTART and all three experiments showed that iSTART increases the extent to which bridging was correlated with comprehension. This latter finding warrants further explication. The effective use of self-explanation requires metacognitive awareness of the strategy and when it is most appropriate (Chi et al., 1989; McNamara & Magliano, 2009a). Research has shown that iSTART training increases the quantity and some aspects of the quality of self-explanations (Magliano et al., 2005; McNamara, 2004; McNamara et al., 2006; McNamara et al., 2009). However, we contend that the current results hint at the possibility that iSTART also improves self-explanation by promoting the metacognitive awareness of when the sub-strategies that comprise self-explanation are most appropriate. That is, the findings indicating that bridging scores were more predictive of comprehension performance after training suggests that students are better able to discern when they are needed to promote comprehension after training than before.

Additionally, iSTART reduced the impact of paraphrasing on comprehension. Although, iSTART teaches students how to paraphrase, it also teaches students that paraphrasing serves primarily as a good starting point for deep comprehension (McNamara, 2004; McNamara et al., 2004). That is, a good paraphrase does not enhance comprehension, but rather allows one to pick out what information needs to be informed by prior text or world knowledge (McNamara & Magliano, 2009a)—information sources drawn upon by bridges and elaborations. It is interesting to note that although there was a significant increase in paraphrasing after training for high school students, this variable was not a significant predictor of comprehension after training. This supports the argument that paraphrasing alone does not support effective self-explanation.

It may appear to be surprising that only skilled readers appeared to benefit from iSTART in terms of increasing the frequency of bridging, but this is the third replication of this finding (Magliano et al., 2005, McNamara et al., 2006). What is notable about the present study is that we replicated that effect using an automated coding of the protocols in the context of RSAT. Although the improvement of bridging for higher ability readers is encouraging, interventions such as these are also intended to help struggling readers. It would be desirable if less skilled readers also benefitted from training. However, iSTART requires considerably more practice than afforded in the version used in this study before less skilled readers benefit from it (Jackson, Boonthum, & McNamara, 2010; Jackson et al., in press). Moreover, it may be the case that training should be tailored to the needs of students, which is the work of future studies. The assessment of comprehension strategy use provided by RSAT could provide a basis for tailoring training. More research is required to determine how best to accomplish this.

It is important to acknowledge that there was no evidence that exposure to iSTART improved comprehension performance, which is not consistent with prior research (e.g., Magliano et al., 2005; McNamara et al., 2006). There is at least one important difference between the comprehension measure used in this study and those used in prior studies. The comprehension questions were embedded in the text and administered in the context of RSAT, whereas prior studies have presented texts and then presented the comprehension test. It may be the case that both of these differences made the comprehension measure provided by RSAT less sensitive to changes in comprehension performance due to exposure to training. Because of this, one may express caution when using the embedded questions as an assessment,assessment; however, previous research has shown that performance on the embedded questions correlates robustly with outcome measures of ability, such as the ACT (r = .54) and Gates-MacGinitie Reading Test (r = .53) (Magliano, et al., 2011).

These results also demonstrate the general utility of RSAT to measure comprehension processes prior to and after training. We see two potential uses of RSAT in the context of iSTART and reading strategy interventions in general. First, RSAT provides information regarding the quality of self-explanations for individual students prior to training, which could be used to tailor training to the needs of individual students. The second is as a formative assessment that could provide teachers and students a basis for assessing strengths and weaknesses in the use of self-explanation after training. One could envision administering RSAT several times after training is completed and providing students with feedback regarding their use of the strategies emphasized during training. It is well documented that learning complex strategies requires sustained and immediate feedback (Kulik, & Kulik, 1988) and RSAT could provide an effective vehicle for doing so after iSTART.

These results also add to a growing body of literature suggesting the viability of computer-based assessments of open-ended student products (Foltz, 2007; Landauer, Laham, & Foltz, 2003; Magliano & Graesser, in press; Magliano & Millis, 2003; Magliano et al., 2002; Millis, Magliano, Todaro, 2006). We have used both sophisticated word counts and LSA to provide such assessments (e.g., Millis et al., 2006). Our approach typically involves using these tools to assess the semantic overlap between the students' protocols and semantic benchmarks. In the case of RSAT, we compare their protocols to the current sentence and prior sentences, which provide measures of paraphrasing and bridging, respectively (the RSAT measure of elaboration is derived from an assessment of new words produced by the students). These assessments are correlated with human judgments of the strategies present in the protocols (Millis et al., 2006).

These data suggest the general approach of estimating processes and strategies based on the content of verbal protocols is sufficient for assessing changes as a function of strategy training. However, more precision in the detection of the use of specific strategies could be of benefit in terms of assessing the use of specific reading strategies before and after training.

In summary, this study suggests that iSTART changes the effectiveness of self-explanations in promoting comprehension. We contend that these changes reflect a better understanding of when particular strategies are most appropriate to promote comprehension. These changes may not be adequately revealed by exploring quantitative changes in the use of strategies after training. Rather researchers, and ultimately the administrators of training in the field (e.g., teachers and reading tutors), need to assess how effectively students use the strategies to accomplish their comprehension goals. Automated protocol assessment tools, such as RSAT, will be vital to providing such assessments.

Acknowledgements

The research and preparation of the manuscript was supported by the Institute for Education Sciences (Grants R305G040055 and R305G040046), partially supported by National Institute of Health (Grants T32 AG000030–31 and RO1-MH70674), and supported by the Center for the Interdisciplinary Study of Language and Literacy (CISLL) at Northern Illinois University. Any opinions, findings, and conclusions or recommendations expressed in this article are those of the authors and do not necessarily reflect the views of the IES, NIH, or CISLL.

References

Bellissens, C., Jeuniaux, P., Duran, N., & McNamara, D. (2007). Towards a textual cohesion model that predicts self-explanations inference generation as a function of text structure and readers' knowledge levels. *Proceedings of the 29th Annual Meeting of the Cognitive Science Society* (pp. 233-238). Austin, TX: Cognitive Science Society.

Birtwisle, M. (2002). The Soundex Algorithm. Retrieved from: http://www.comp.leeds.ac.uk/matthewb/ar32/basic_soundex.htm.

Chi, M.T.H., Bassok, M., Lewis, M.W., Reimann, R., & Glaser, R. (1989). Self-explanation: 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.

Coté, N., & Goldman, S.R. (1999). Building representations of informational text: Evidence from children's think-aloud protocols. In H. van Oostendorp & S.R. Goldman (Eds.), *The construction of mental representations during reading* (pp. 169-193). Mahwah, NJ: Lawrence Erlbaum Associates.

Coté, N., Goldman, S.R., & Saul, E.U. (1998). Students making sense of informational text: Relations between processing and representation. *Discourse Processes*, *25*, 1-53.

Christian. P. (1998). Soundex – can it be improved? *Computers in Genealogy, 6*, 215-221.

de Leeuw, N., & Chi, M.T.H. (2003). The role of self-explanation in conceptual change learning. In G. Sinatra & P. Pintrich (Eds.), *Intentional Conceptual Change* (pp. 55-78). Mahwah, NJ, US: Lawrence Erlbaum Associates Publishers.

Foltz, P.W. (2007). Discourse coherence and LSA. In Landauer, Thomas K., McNamara, Danielle S., Dennis, Simon & Kintsch, Walter (Eds.), *Handbook of latent semantic analysis* (pp. 167-184). Mahwah, NJ, US: Lawrence Erlbaum Associates Publishers.

Gilliam, S., Magliano, J. P., Millis, K. K., Levinstein, I., & Boonthum, C. (2007). Assessing the format of the presentation of text in developing a Reading Strategy Assessment Tool (RSAT). *Behavior Research Methods, Instruments, & Computers*, *39*, 199-204.

Graesser, A. C., Singer, M., & Trabasso, T. (1994). Constructing inferences during narrative text comprehension. *Psychological Review, 101*, 371-395.

Jackson, G.T., Boonthum, C., & McNamara, D.S. (2010). The efficacy of iSTART extended practice: Low ability students catch up. In J. Kay & V. Aleven (Eds.), *Proceedings of the 10th International Conference on Intelligent Tutoring Systems* (pp. 349-351). Berling/Heidelberg: Springer.

Jackson, G. T., Demspey, K. B., & McNamara, D. S. (in press). Game-based practice in a reading strategy tutoring system: Showdown in iSTART-ME. In H. Reinders (Ed.), *Computer games*. Bristol, UK: Multilingual Matters.

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

Kulik, J. A., & Kulik, C-L. C. (1988). Timing of feedback and verbal learning. *Review of Educational Research, 58*, 79-97.

Landauer, T.K., Laham, D., & Foltz, P.W. (2003). Automated scoring and annotation of essays with the Intelligent Essay Assessor. In Shermis, Mark D. & Burstein, Jill (Eds.), *Automated essay scoring: A cross-disciplinary perspective* (pp. 87-112). Mahwah, NJ, US: Lawrence Erlbaum Associates Publishers.

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*, 224–232.

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, and Cognition, 20*, 1456-1470.

MacGinitie, W. H., & MacGinitie, R. K. (1989). *Gates-MacGinitie Reading Test* (3rd ed.). Itasca, IL: The Riverside Publishing Company.

Magliano, J. P., & Graesser, A. C. (in press). Computer-based assessment of student-constructed responses. *Behavioral Research Methods*.

Magliano, J. P., & Millis, K. K. (2003). Assessing reading skill with a think-aloud procedure and latent semantic analysis. *Cognition and Instruction, 21*, 251-284.

Magliano, J. P., Millis, K. K., & Levenstein, I. (2010). Assessing comprehension skill with verbal protocols. Final report submitted to the United States Department of Education for Grant No. R305G040055

Magliano, J.P., Millis, K.K., The RSAT Development Team, Levinstein, I., & Boonthum, C. (2011in press). Assessing Comprehension During Reading with the Reading Strategy Assessment Tool (RSAT). *Metacognition and Learning*.

Magliano, J. P., Wiemer-Hastings, K., Millis, K. K., Muñoz, B. D., & McNamara, D. S. (2002). Using latent semantic analysis to assess reader strategies. *Behavior Research Methods, Instruments, & Computers*, *34*, 181-188.

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.

Magliano, J. P., Trabasso, T., & Graesser, A. C. (1999). Strategic processing during comprehension. *Journal of Educational Psychology, 9*, 615-629.

McNamara, D.S. (2004). SERT: Self-explanation reading training. *Discourse Processes, 38*, 1-30.

McNamara, D.S. (2009). The importance of teaching reading strategies. *Perspectives on Language and Literacy,* 34-40.

McNamara, D.S., Boonthum, C., Kurby, C.A., Magliano, J., Pillarisetti, S., & Bellissens, C. (2009). Interactive paraphrase training: The development and testing of an iSTART module. In the *Proceeding of the International Conference on Artificial Intelligence in Education* (AIED 2009), Brighton, UK, July 6 – 10, 2009.

McNamara, D.S., Levinstein, I.B., & Boonthum, C. (2004). iSTART: Interactive strategy trainer for active reading and thinking. *Behavioral Research Methods, Instruments, & Computers, 36,* 222-233.

McNamara, D.S., Boonthum, C., Levinstein, I.B., & Millis, K. (2007). Evaluating self-explanations in iSTART: comparing word-based and LSA algorithms. In T. Landauer, D.S. McNamara, S. Dennis, & W. Kintsch (Eds.), *Handbook of Latent Semantic Analysis* (pp. 227-241). Mahwah, NJ: Erlbaum.

McNamara, D. S., & Magliano, J. P. (2009a). Self-explanation and metacognition: The dynamics of reading. In D. J. Hacker, J. Dunlosky, & A. C. Graesser, (Eds.), *Handbook of Metacognition in Education*, (pp. 60-81). Mahwah, NJ: Lawrence Erlbaum and Associates.

McNamara, D.S., & Magliano, J.P. (2009b). Towards a comprehensive model of comprehension. In B. Ross (Ed.), *The psychology of learning and motivation*. New York, NY: Elsevier Science.

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

McNamara, D.S., O'Reilly, T., Best, R., & Ozuru, Y. (2006). Improving adolescent students' reading comprehension with iSTART. *Journal of Educational Computing Research, 34,* 147-171.

McNamara, D.S., O'Reilly, T., Rowe, M., Boonthum, C., & Levinstein, I.B. (2007). iSTART: A web-based tutor that teaches self-explanation and metacognitive reading strategies. In D.S. McNamara (Ed.), *Reading comprehension strategies: Theories, interventions, and technologies* (pp. 397-421) Mahwah, NJ: Erlbaum.

McNamara, D.S., & Scott, J.L. (1999). Training reading strategies. *Proceedings of the Twenty-first Annual Meeting of the Cognitive Science Society* (pp. 387-392). Hillsdale, NJ: Erlbaum.

Millis, K. K., & Magliano, J. P. (in press). Assessing comprehension processes during reading. To appear in J. P., Sabatini, J., E. R. Albro, & T. O'Reilly, (Eds.) *Assessing reading in the 21st century: Aligning and applying advances in the reading and measurement sciences*. Lanham, MD: R & L Publishing.

Millis, K. K., Magliano, J. P., & Todaro, S. (2006). Measuring discourse-level processes with verbal protocols and latent semantic analysis. *Scientific Studies of Reading*, *10*, 251-283.

Oakhill, J., & Yuill, N. (1996). Higher order factors in comprehension disability: Processes and remediation. In Cornoldi, Cesare & Oakhill, Jane (Eds.), *Reading comprehension difficulties: Processes and intervention* (pp. 69-91), Mahwah, NJ, US: Lawrence Erlbaum Associates Publishers.

O'Reilly, T., 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 Cognitive Science Society* (pp. 1053-1058). Mahwah, NJ: Erlbaum.

O'Reilly, T., Taylor, R.S., & McNamara, D.S. (2006). Classroom based reading strategy training: Self-explanation vs. reading control. In R. Sun & N. Miyake (Eds.), *Proceedings of the 28th Annual Conference of the Cognitive Science Society* (pp. 1887). Mahwah, NJ: Erlbaum.

Ozuru, Y., Briner, S., Best, R., & McNamara, D. (2010). Contributions of self-explanation to comprehension of high- and low-cohesion texts. *Discourse Processes, 47,* 641-667.

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

Spilich, G.J., Vesonder, G.T., Chiesi, H.L., & Voss, J.F. (1979). Text processing of domain-related information for individuals with high and low domain knowledge. *Journal of Verbal Learning & Verbal Behavior, 18*, 275-290.

Trabasso, T., & Magliano, J. P. (1996).

Table 1.

Example indirect protocols for the sentence "Cancer may take many years to accumulate.," from the text "How cancer develops"

Participant	Protocol	Current Sentence Word Matches	Paraphrase Score	Local Text Word Matches	Local Bridging Score	Distal Text Word Matches	Distal Bridging Score	Mentioned Words Not in Text So Far	Elaboration Score
1	Three to seven mutations in a cell are required to cause cancer. It can take a long period of time to develop.	cancer, take	2	cell, mutations, seven, three	4	develop	1	required, period	2
2	Cancer may take many years to develop follows the sentence that tells me that 3 to 7 mutations must occur before a cell becomes cancerous. These mutations (or the number of mutations that it takes to make a cell cancerous) may take many years to accumulate and develop.	accumulate, cancer, take, cancerous, years, takes	6	becomes, cell, mutations, occur	4	develop, make	2	follows, number, tells, sentence	4
3	The progression of cancer. It can happen over such a short period of time and the growth rates of different cancers.	cancer, cancers	2	different	1	NA	0	short, rates, progression, happen, growth, period	6
4	How many people have signs of cancer and may not know if yet, because of how long it can take to develop.	cancer, take	2	NA	0	develop	1	know, signs, people	3
5	Cancer is very harsh on the DNA.	cancer	1	NA	0	dna	1	harsh	1

Note: The paraphrase score represents the number of word matches in the protocol from the current sentence. The local bridging score represents the number of word matches from the immediately prior sentence. The distal bridging score represents the number of word matches from all previous sentences except the immediately prior sentence. The elaboration score represents the number of content words present in the protocol that were not mentioned in the prior text or current sentence.

Table 2.

Example answers to the question, "Why would this happen?" after reading the sentence "The cell may then divide too often.," from the text "How cancer develops."

Participant	Protocol	Ideal Answer Word Matches	Direct Question Score
1	A mutations in the cell will cause the rapid growth.	growth, cell, mutations	3
2	In cancerous cells, the genes within the nucleus malfunction; they can allow a cell to divide and divide without incoming messages to stimulate it each time. The growth factor is increased greatly.	growth, cell, stimulate, cells, factor, genes, divide	7
3	Because the proto-onconogenes are giving out receptors that are not natural to the human body and telling them to divide more rapidly.	proto, divide	3
4	Because it just divides on its own, it doesn't get instructions from any other cell.	cell, instructions, divides	3
5	Because there are to many genes.	genes	1

Note: The ideal answer is "The proto-oncogene mutates into an oncogene-a gene that instructs the cell to grow and divide repeatedly without stimulation from neighboring cells. Some oncogenes overproduce growth factors." The direct question score is represented by the number of word matches between the answer provided by participants and the ideal answer.

Table 3.

Mean RSAT scores for pre- and post-training session as a function of strategy training.

Measure	Condition	Session	
		Pre	Post
Paraphrasing	iSTART	1.05 (0.07)	0.88 (0.05)
	Control	0.85 (0.07)	0.87 (0.05)
Local Bridging	iSTART	0.66 (0.06)	0.54 (0.04)
	Control	0.53 (0.06)	0.56 (0.04)
Distal Bridging	iSTART	1.71 (0.11)	1.34 (0.08)
	Control	1.52 (0.12)	1.20 (0.08)
Elaboration	iSTART	4.06 (0.26)	3.49 (0.18)
	Control	3.91 (0.26)	3.50 (0.19)
Comprehension	iSTART	2.25 (0.11)	1.98 (0.11)
	Control	2.10 (0.11)	1.99 (0.11)

Note: Standard errors are reported in the parentheses.

Table 4.

Predicting performance on RSAT comprehension score prior to and post iSTART training.

Predictor variables		iSTART		Control	
		Pre-Training	Post-Training	Pre-Training	Post-Training
Paraphrase score	*Beta*	0.76	0.12	0.06	0.29
	SE	0.35	0.24	0.28	0.44
	t	3.16	0.85	0.37	1.84
	p	0.003	0.401	0.712	0.072
Bridging score	*Beta*	-0.25	0.42	0.39	0.33
	SE	0.17	0.12	0.15	0.21
	t	.99	2.91	2.25	1.82
	p	0.320	0.011	0.030	0.077
Elaboration score	*Beta*	0.29	0.23	0.19	0.9
	SE	0.07	0.07	0.08	0.07
	t	2.09	1.91	1.19	0.60
	p	0.043	0.063	0.240	0.550
ACT scores	*Beta*	0.28	0.28	0.27	0.06
	SE	0.03	0.02	0.03	0.03
	t	2.38	3.08	2.25	0.46
	p	0.022	0.004	0.042	0.643
Estimated Variance Explained		46%	61%	44%	35%

Table 5.

Mean RSAT scores for pre- and post-training session as a function of strategy training and skill.

Paraphrasing

Training	Skill	Session	
		Pre	Post
iSTART	Skilled	1.16 (0.11)	1.37 (0.11)
	Less Skilled	1.65 (0.11)	1.93 (0.11)
Control	Skilled	1.22 (0.12)	1.25 (0.12)
	Less Skilled	1.73 (0.11)	1.80 (0.11)

Bridging

Training	Skill	Session	
		Pre	Post
iSTART	Skilled	2.68 (0.11)	3.44 (0.24)
	Less Skilled	1.77 (0.21)	1.81 (0.23)
Control	Skilled	3.03 (0.21)	3.07 (0.23)
	Less Skilled	1.84 (0.23)	1.95 (0.25)

Elaboration

Training	Skill	Session	
		Pre	Post
iSTART	Skilled	5.65 (0.49)	4.99 (0.37)
	Less Skilled	4.20 (0.49)	3.55 (0.37)
Control	Skilled	5.90 (.50)	3.65 (0.38)
	Less Skilled	4.69 (0.54)	4.16 (0.41)

Note: Standard errors are shown in parentheses

Table 6.

Predicting performance on RSAT comprehension score prior to and post iSTART training.

Predictor variables		iSTART		Control	
		Pre-Training	Post-Training	Pre-Training	Post-Training
Paraphrase score	*Beta*	0.50	-0.19	-0.002	-0.18
	SE	0.18	0.21	0.20	0.19
	t	2.92	0.95	0.01	1.00
	p	0.005	0.345	0.992	0.321
Bridging score	*Beta*	007	0.71	0.45	0.66
	SE	0.10	0.10	0.11	0.09
	t	0.38	3.33	2.24	3.78
	p	0.707	0.001	0.028	0.001
Elaboration score	*Beta*	0.30	0.13	-0.01	0.01
	SE	0.03	0.10	0.028	0.05
	t	2.99	1.14	0.07	0.12
	p	0.004	0.245	0.949	0.909
Estimated Variance Explained		43%	40%	20%	29%

Table 7.

Means and standard deviations for RSAT scores for college and high school students.

Predictor variables	High School Students	
	Pre-Training	Post-Training
Paraphrase score	1.00 (0.59)	1.25 (0.71)
Bridging score	1.34 (0.90)	1.47 (1.02)
Elaboration score	3.17 (1.72)	3.16 (1.50)
Comprehension score	1.56 (0.78)	1.73 (0.87)

Table Note: Standard errors are presented in parentheses.

Table 8.

Predicting performance on RSAT comprehension score prior to and post iSTART training.

Predictor variables		High School Students	
		Pre-Training	Post-Training
Paraphrase score	*Beta*	0.43	0.15
	SE	0.24	0.22
	t	2.39	0.83
	p	0.02	0.41
Bridging score	*Beta*	0.16	0.37
	SE	0.18	0.15
	t	0.76	2.13
	p	0.45	0.04
Elaboration score	*Beta*	0.15	-0.07
	SE	0.07	0.11
	t	0.96	-0.39
	p	0.34	0.70
Gates	*Beta*	0.15	0.37
	SE	0.68	0.86
	t	0.96	2.05
	p	0.35	0.05
Estimated Variance Explained		40%	47%