Reading and Writing (2005) 18:755–786
DOI 10.1007/s11145-005-0956-5

Web-based training of metacognitive strategies for text comprehension: Focus on poor comprehenders

MINA C. JOHNSON-GLENBERG
Waisman Center University of Wisconsin Madison, WI, USA; The NeuronFarm, LLC Madison, WI, USA

Abstract. Metacognitive reading strategies were trained and practiced using interactive Web-based tools. Twenty middle school poor reading comprehenders were trained in two metacognitive strategies using a Web-based application called 3D-Readers. The training texts were science-oriented and merged the narrative and expository genres. Results from a within-subjects design answered two main experimental questions: (1) Were greater comprehension gains demonstrated after reading experimental texts with embedded verbal (generate questions) and visual (create a model) strategies compared to control texts? (2) Did the embedded strategies affect elective rereading of the texts? The data answered both questions in the affirmative. Comprehension, as assessed with constructed answers, was significantly higher in the experimental condition, thus demonstrating the efficacy of training verbal and visual strategies in a Web-based environment. In addition, participants elected to reread more often in the experimental condition (as assessed with number of clicks to "ScrollBack" through the text), thus demonstrating the efficacy of strategy training on text reprocessing. Interestingly, the *poorer* comprehenders altered their rereading behavior the most. Implications for Web-based instructional applications are discussed.

Key words: At-risk readers, Comprehension, Computer-assisted learning, Metacognitive strategies, Reading remediation, Struggling readers, Web-based applications

It is estimated that 95% of schools in the US have Internet access for learning purposes (http://www.nces.ed.gov). Many classroom teachers are turning to the Web for instructional purposes and there is great potential for Web-based training of higher level cognitive activities like reading comprehension. However, new learning programs must be designed using research-based educational theory to take advantage of the interactivity and synchronicity of the Web. Students can now get immediate feedback and guidance regarding free text performance, i.e., it is now possible to move beyond the somewhat rigid assessment formats of multiple choice and automatically score constructed text on-line. Learning module designers need to be aware of "potential mismatches between current technologies and human learning'' (Reyna, Brainerd, Effken, Bootzin, & Lloyd, 2001, p. 40). This means that design elements should be rigorously researched, analyses should include aptitude by treatment interactions, and elements should be tested on heterogeneous non-mainstream populations. This study focuses on poor reading comprehenders and whether better reading practices can be trained using Web-based tools.

The technology used in this study is based on an experimental software program called 3D-Readers that harnesses the immediacy and dynamicism of the Web to teach, assess, and allow practice on two powerful metacognitve reading strategies, one verbal and one visual. The verbal strategy involves the creation of questions while reading the text, and the visual strategy involves the facilitation of building a mental model by manipulating images on the screen that construct main ideas. The texts are hybrids of both the expository (science) and narrative (prosodic) genres. Narrative is a natural category and is poorly served by sharp boundary distinctions. Graesser, Golding, and Long's (1991) description is elegant, ''Narratives are expressions of event-based experiences that (a) are either stored in memory or cognitively constructed, (b) are selected by the writer to transmit to the reader, (c) are organized in knowledge structures that can be anticipated by the audience'' (p. 174). This study's hybrid texts were created to facilitate comprehension in struggling readers. The working hypothesis was that using a familiar knowledge structure would decrease the cognitive load associated with processing the novel scientific concepts. Expository text is more difficult to understand and retain. Narrative resembles more closely orality. The knowledge of narrative story structure is acquired before reaching school age (Stein & Glenn, 1979). A further benefit from narrative may be that the genre depicts ''event sequences that people in a culture directly enact or experience'' (Graesser, Golding, & Long, 1991, p. 172).

To better understand the training program, a quick synopsis of a typical session is presented. Each time readers log on to an experimental text on the Website, the following order of events occurs:

(a) Readers first indicate whether they have prior knowledge of the topic. This task activates prior knowledge and facilitates comprehension. (Because the prior knowledge assessment in this first iteration involved only a coarse-grained yes/no answer and did not correlate with any other variables, it will not be discussed further.)
(b) Readers answer seven vocabulary questions, which contain some of the more difficult words from the text. These answers are used in conjunction with post-reading vocabulary to assess changes in knowledge. They further activate prior knowledge.

(c) Readers read a section of text and at an appropriate moment are prompted to work on a *verbal strategy*, i.e., they create a question. They see and hear the instructions "Pretend you are a teacher. What question would you ask here?" Readers type in the question and it is sent to an automatic scoring algorithm.
(d) After the next text section, readers work on a *visual strategy* designed to facilitate imagery skills and deepen understanding. The visual strategies are colorful screens split in two. Typically, on the right side is the toolbox filled with icons depicting important conceptual elements from the text, and on the left is the "model space" where the icons are assembled. That is, readers drag icons into the model space to create a visual representation of one of the main ideas.
(e) Readers engage in one more verbal strategy, and one more visual strategy for a total of four strategies. Note that before each strategy they are prompted to reread if there was anything they did not understand. They then exit the text and are not able to re-read.
(f) Readers answer the seven vocabulary questions again as a post-test.
(g) Finally, readers answer eight open-ended comprehension questions. Their constructed answers are scored immediately by the automatic scoring algorithm. Readers are shown their average score after the eight questions so that they can track their progress over time.

The strategies

According to Gough and Tunmer (1986) reading can be portioned into two major components: decoding and linguistic comprehension. This means that if readers have adequate decoding skills (and motivation), but still have problems understanding what they have read, then their primary deficit lies at the higher level of linguistic comprehension. This would also include the interfacing between the two skill levels. How can the power of the Web be used to remediate higher level reading comprehension deficits? Automatic essay scoring algorithms can be used to score generated questions and answers to questions, and advances in streaming technology now allow for colorful, interactive graphics that might facilitate the construction of mental models. Equally important is the aspect of immediate, personalized feedback so that readers can repair miscomprehension while still engaged with the text.

The meta-analysis report of the National Reading Panel in America (NRP, 2000), lists 16 categories of comprehension instruction, including mental imagery. The NRP results section stipulates that seven of these categories demonstrate *firm scientific evidence* supporting their efficacy in improving comprehension in typical readers. This study's Web-based instructional program utilizes five of these seven strategies: Comprehension monitoring, graphic organizers, question answering, question generation, and summarization. (Some summarization is always required in the final answers). This study's training package does not use semantic organizers or cooperative learning, although there are plans to reconfigure it for the latter.

The two trained strategies are considered metacognitve. This study operates from a definition put forth by Paris, Cross, and Lipson (1984). Metacognition has two fundamental aspects: knowledge about cognition and self-directed thinking. Self-directed thinking is governed by evaluation, planning, and regulation activities. This program aims to modify these three activities. The hypothesis is that with repeated practice on the verbal and visual strategies, and with repeated immediate feedback on performance, readers will be aided in their *evaluation* skills. Evaluation will, in turn, help activate *planning* and *regulation* activities. With practice, this evaluation – or monitoring and self-assessment – should become more automatic. Poor comprehenders need fast, flexible, internal, and tangible strategies that they can call on in times of miscomprehension. A mixture of both verbal and visual metacognitive strategies was chosen for the first iteration of study.

Johnson-Glenberg (2000) describes a metacognitve intervention study with 45 adequate decoders, but poor comprehenders. In that study 3rd through 5th graders (average chronological age 9–11) were randomly assigned to three groups. Two groups were instructed in two very different comprehension strategies. The first was the package of verbal strategies from Reciprocal Teaching (RT) created by Palincsar and Brown (1984), and the second was a visual package called Visualizing/Verbalizing (V/V) created by Bell (1986, 1991). The third group was composed of untreated controls. RT is a comprehension-monitoring and comprehension-fostering package containing four strategies: summarization, prediction, clarification, and question generation. V/V is a primarily visual strategy in which readers learn to mentally create, and then describe "movies in their heads" as they read. After 10 weeks of training, the two experimental groups demonstrated significant gains on several measures related to reading comprehension when compared to the untreated control group. The experimental groups constructed better answers on both implicit and explicit questions, and demonstrated gains in word recognition skills. The research supported the hypothesis that strategy training aided the readers, and that *both verbal and visual strategies were effective*.

Some strands of research, and many classroom teachers, refer to these strategies as visual. Although it may be more accurate to call the strategies imaginal, we would like to be consistent in our publication terminology and will refer to them as visual. This study is multicomponential. In part because research supports that comprehension instruction should contain multiple components (National Reading Panel, 2000; Pressley, 2000), and in part because resources did not allow for a mixed 2×2 crossing the verbal and the visual strategies.

The verbal strategy – creating questions. Of all the specific verbal strategies available, question generation is supported by the strongest scientific evidence (National Reading Panel, 2000). Rosenshine, Meister, and Chapman (1996) reviewed intervention studies that taught students how to generate questions and found that training yielded an impressive median post-training effect size of .86 on experimenter-designed measures.[1] As further support, in a 1996 study, Trabasso and Magliano gave third grade students three stories to read. After the second story, one group of students was asked "why" and "how" questions. This group recalled significantly more of the third story. In addition, the "how" and "why" question group gave qualitatively more sophisticated explanations during the later think-aloud protocol. These results suggest that after being asked "how" and "why" questions subsequent comprehension is enhanced because readers must generate causal links and goal plans. The current study takes this construct a step further, and allows the readers to ask the "how" and "why" questions of themselves. Tovani (2000) asserts that readers who ask questions improve their comprehension by interacting with the text and remaining focused on it. Asking questions fosters curiosity because readers are forced to continue reading to answer their questions, clarify information in the text, and, finally, go beyond literal meanings and to begin the first step in the processes of deduction or inference.

At appropriate locations in the text, readers are prompted for questions and they type in their questions. These are automatically scored for quality. Quality was assessed with several methods to be described later. The two main criteria for higher scores were that the *answer* be multiword and that the question began with "how" or "why". For the first criterion, the answer to the question should require a sentence or many words to answer. Effective questions were ones that required deeper processing of the text or even inferencing. Examples were given of questions that were: not very effective – "Did you like the story?", more effective/good – "What part of the eye do you see when someone has 'red-eye' in a photograph?", and most effective/excellent – "How does someone get 'red-eye' in a photograph?". For the second criterion, it was explained to the readers before the experimental condition that generally questions that started with either "how" or "why" were requesting information that required deeper processing.

The visual strategy – building mental models. The ultimate act of reading is the creation of a mental model (Johnson-Laird, 1983 – also called a situation model, Kintsch, 1988). Although not all mental models are visual, research on the importance of imagery and visualization during information processing and reading has a long and respected history. Based primarily on Paivio's Dual Code Theory (Paivio, 1986; Sadoski & Paivio, 2001), numerous studies report significant comprehension gains by experimental groups that were either encouraged, or taught to visualize while reading (Bell, 1991; Gambrell & Jaywitz, 1993; Johnson-Glenberg, 2000; Levin, 1973; Mayer & Sims, 1994; Oakhill & Patel, 1991; Pressley, 1977). Creating a visual model on the screen concretely aids poor readers in building internal visual models. By manipulating and building images, readers discover and/or confirm how the segments of sequential text fit together in a 'three dimensional' gestalt. In this current study an incomplete, stylized version of a visual "mental model" appears on the screen. The reader interacts with the model, and is in turn supplied with immediate feedback regarding performance.

The visual strategies are designed to serve two purposes: (1) to *train* in comprehension by serving as a stepping stone between text and mental representation, and (2) to *assess* text comprehension. As an example, readers might first be shown an animation of how a practice sentence should be built (the training stage). The practice sentence at the top of the screen might explain that the color of an object is really the lightwaves that the object is reflecting. Therefore, a banana is reflecting…? The animation would show the correct icons being dragged out of the toolbox and placed in order – first a flashlight which would emit white light, and then a banana, and then rays with a yellow hue that bounce off of the banana. Readers are then expected to build their own model of a real sentence from the text which would require near transfer (e.g., "What color would the iguana be reflecting?"). Figure 1 illustrates one black and white screenshot of this intensely colorful interactive visual strategy.

When readers feel confident about their self-assembled models, they submit them for scoring and receive a percent correct score (the assessment stage). Readers are allowed up to three submissions. After the final submission, if the reader does not have a 100% score, the system uses animation to build the correct model.

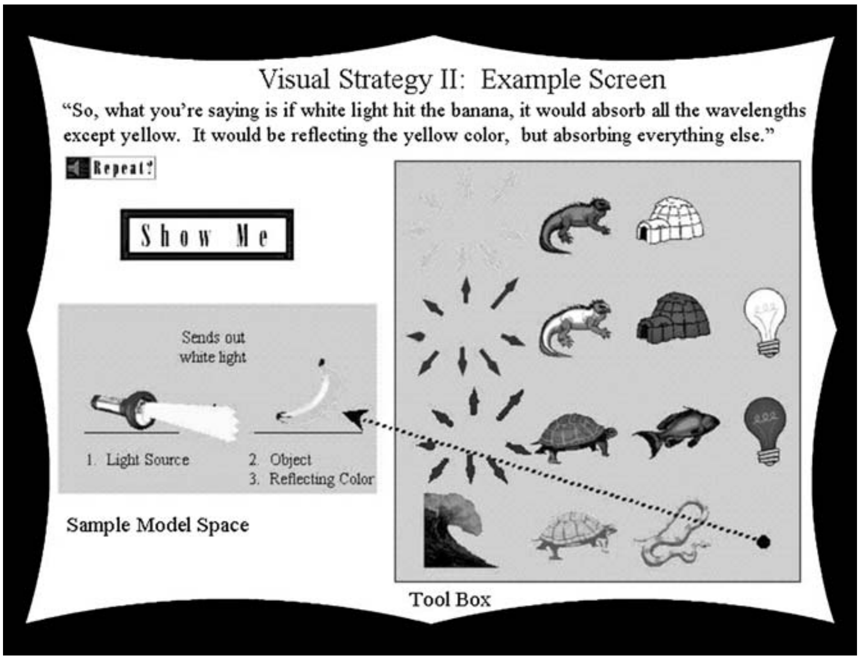

Figure 1. Screenshot of an interactive visual strategy.

Rereading (ScrollBacks). Rereading has been operationalized as ScrollBacks. Readers could click on one of two buttons to scroll back through a section of text to reread it. Only four lines of text were visible at a time. Mature (good) readers use the strategy of rereading (Pressley, 2000). Memory for text has been shown to improve with rereading (Amlund, Kardash, & Kulhavy, 1986). Millis, Simon, and tenBroek (1998) posit that rereading facilitates comprehension because it allows readers to complete processes which produced less than ideal outputs from prior reading. Poor readers are often disinclined to reread text. They may not understand that the goal of reading is to extract meaning (several poor readers have reported to the author that the goal of reading is to "get to the end"). Poor comprehenders may be unaware that they are not comprehending, or they may not know that such a strategy even exists. We included rereading as a variable in this study to ascertain whether introduction to and practice on other higher-order, more metacognitive reading strategies might affect the strategy of rereading.

Vocabulary. The pre-reading vocabulary test serves as a brief assessment of content knowledge and stimulates prior knowledge. Seven of the more difficult words are culled from the text, and a multiple choice format is used for assessment. Although multiple choice has its limitations, it *is* quickly administered and allows for a straight-forward subtraction method to assess change in knowledge at the time of post-test. See Appendix A for a sample of the vocabulary test.

A word on essay scoring algorithms

Most reading assessment relies on answering multiple choice questions. This is a limited knowledge assessment paradigm. In addition to adding a level of chance (usually 25–33%) to a reader's score, it relies on a rigid, one-correct-answer template. This format erroneously sends readers in search of the "one true meaning" of the text (Pearson & Hamm, 2001). The eight final questions in this study are open-ended and require students to construct more authentic answers. The system uses the High-dimensional Expert Match Algorithm (HEMA) to score a reader's generated questions and constructed answers.

The inclusion of HEMA also means that there are checks and balances in the system. One of the attractive elements of employing both visual and verbal strategies is that the system can immediately cross-reference a reader's performance on both types of strategies and alert teachers to any discrepancies. This could happen in either the verbal or visual domains. It is important to note that even though HEMA represents the vanguard of computerized automated scoring, a computer will *never* be as flexible and creative as a human scorer. No claims are made that the processing in HEMA is isomorphic to human cognition. HEMA is merely a tool designed to mimic, as closely as possible, human scoring of written language. Language is by definition generative and evolving; decades of AI research have taught us that the true correlation between human expert assessment and computer-generated assessment will never be 1.00. HEMA makes this Web-based remediation system unique in that it allows readers to create their own questions (which in the next iteration will be scored and sent as feedback). HEMA allows readers to answer open-ended questions and receive immediate feedback. Space constraints do not allow for more detail on the algorithm; it is mentioned here primarily as a reliability measure for the human scorers.

Predictions

To measure the outcome validity of the strategy training, two different comprehension measures were chosen as dependent variables. The first was constructed answers, the prediction was that in the experimental condition (with the embedded strategies) participants would demonstrate significantly greater gains on the final answers – as scored by both human experts and the HEMA algorithm. The second variable was vocabulary test gain from pretest to post-test, the prediction was that in the experimental condition the gain would also be greater. Data were gathered on number of ScrollBacks (rereads), and the scoring of the readers' generated questions (by both humans and HEMA). The prediction was that in the experimental condition readers would use the ScrollBack mechanism more frequently because they would be engaged in deeper text processing and would now be more aware or cognizant of the fact that they needed to reread a section. It was also predicted that readers who generated higher level questions would also demonstrate better comprehension on the final constructed answers.

Method

Participants

The study included 20 participants from one middle school in a lower-middle class neighborhood of a small midwestern city in the US. Six participants were in the 7th grade and 14 were in the 6th grade with an average chronological age of 12.5 (range from 11.10 to 13.6). The reading specialist at the school handled the majority of recruiting and scheduling, utilizing prior contact with the students and/or referrals from teachers. We asked teachers for students who were poor comprehenders. We specified that they either be below the mean on the state's standardized reading test (TerraNova Test – CTB/McGraw-Hill, 1997) or currently reading below the mean for the class in text comprehension. We reminded teachers that these children are often poor listeners as well, so they might also be the children who do not follow directions well and ask repeatedly, "What are we supposed to be doing?" In addition, we stipulated that they be able to decode texts written for the middle of 5th grade. Poor comprehenders were selected for two reasons. First, they are the primary target audience for this software, and second, they are the readers who often show the greatest gains from comprehension instruction programs. The poor comprehenders in Oakhill and Patel's (1991) study demonstrated a greater benefit from imagery training than the better comprehenders. Furthermore, teachers were asked to not recommend readers who were English language learners (ELL) or who were classified as having special needs (EEN-Educable Exceptional Needs).

Standardized reading scores from the previous spring's TerraNova Test (CTB/McGraw-Hill, 1997) which measures both decoding and comprehension, revealed an average percentile score of 33.26% (SD = 17.30) for the sample. One student was above the 50%ile (at the 72%ile), which was surprising since he was only the 8th ranked in final comprehension scores in this study. Perhaps his teacher felt that he was starting to lag behind his classmates during the fall. Because deleting his data did not alter the results, his data have been retained in the analyses. Two participants left the study. One was a new transfer student who apparently was decoding at around the 3rd grade level, and one self-selected to stop after several sessions so that he could remain in his classroom. The ethnicity breakdown was as follows: 45% African-American, 40% White, 10% Asian, and 5% Hispanic. The sample was 55% female.

Materials

Room and computers. The main room for the study was the computer lab. It was a glassed-in room in the middle of the library. When all students were present in the two largest groups, two computers were used in the library and a trainer stayed in the library with them. There were rarely other students in the library at this time. The computers were all Internet connected PCs, of 486 speed or higher; the majority manufactured by Compaq.

Texts. The seven original texts (the practice text was repeated once) were written by the author and two middle school teachers whose specialities were science and reading instruction. The texts were rewritten on average six times by committee until they were balanced on several key variables. Very specific guidelines were created and followed, the texts had to be (a) engaging in an age-appropriate and prosodic manner, (b) synchronized with the State of Wisconsin middle school content standards for science (via FOSS – Full Option Science Program, Lawrence Hall of Science; http://www.lhs.berkeley.edu/foss/), (c) visual enough to facilitate the graphical mental model strategies, and (d) less than 1,800 words so they could be finished in the half hour time period. The topics, which were appropriate for the school district's content standards, were (1) how to measure a wave (for the short practice texts), (2) volcanoes, (3) satellites, (4) color as wavelength, (5) how the eye works, (6) biomes, and (7) telescopes. The texts contained an average of 1,445 words and the average Flesch-Kincaid Readability score was 5.22 (beginning of 5th grade). The texts were written with a mixture of narrative and expository elements. For example, in the text on biomes the twin brother and sister hiked up a mountain where they experienced the four major biomes in one setting – as they joked/fretted about being followed by a bear. See Appendix B for the text on color as wavelength.

ScrollBacks. We operationalized rereading with ScrollBacks. Text was double spaced 14 point font. In a window in the center of the screen four lines (including blanks) were legible at a time. The text before and after the reading window was masked with case-sensitive Xs and the background was grey. The reading text background was white. On the right side of the screen were four buttons. The bottom two buttons scrolled the reader forward, and the top two scrolled the reader back through one section of text. There were five sections per text. The double upward-pointing arrow scrolled readers up five lines of text. The single upward-pointing arrow scrolled readers up one line of text. This was explained to readers in the first session. Figure 2 illustrates a text page and the scrolling interface.

HEMA. Because we were committed to assessing knowledge through constructed, or free text, we needed to design an algorithm powerful, yet

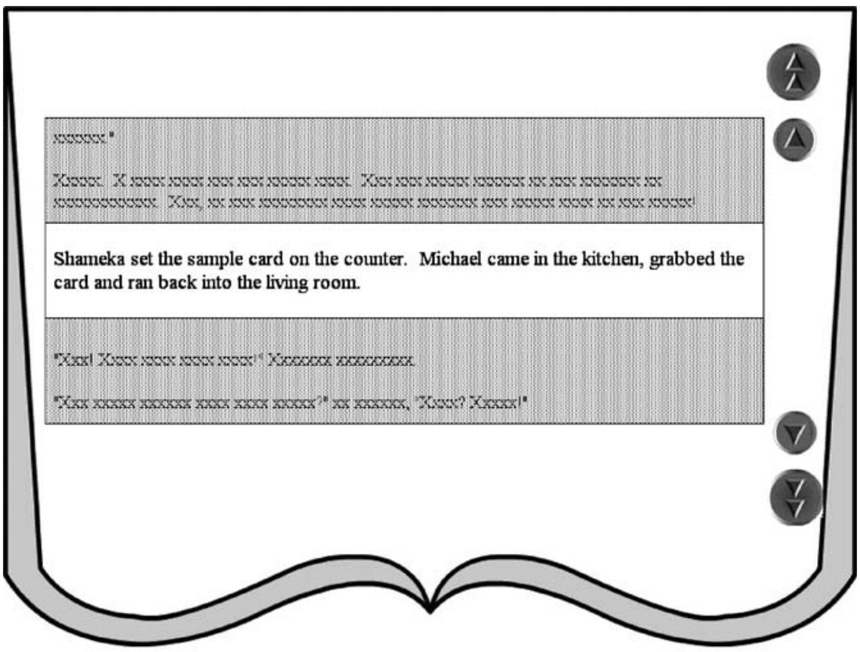

Figure 2. Example of the text reading and scrolling interface.

flexible, enough to score the range of readers' responses. Among several versions of automatic essay scoring algorithms on the market, one of the leaders is KAT (Knowledge Assessment Technologies) based on Latent Semantic Analysis (Landauer & Dumais, 1997). The algebra in HEMA is significantly different from other automated systems and sometimes includes neural networks. Our purpose is different as well, it is to assess the very short, very specific types of question and answers that children ages 10 through 17 create for hybrid-style texts. We designed two versions of HEMA to be able to score both the questions generated and the answers constructed. For purposes of this article HEMA is only discussed insofar as it supplements interrater reliability.

Design

This study was allowed access to only one school and because it was a ''pull-out'' program with a specific population, it had access to a small number of participants. To control for variability a within subjects design was utilized. Because it was necessary to also control for cognitive carry-over effects associated with educational interventions, the control condition always came first. In the control condition participants unscrambled four anagrams embedded in the texts. In the experimental condition the two metacognitive strategies were embedded twice in the texts (for a total of four). To control for experimental story effects, order of text was varied between four order conditions in a modified Latin Square. Participants were randomly assigned to order condition. Participants were seen in three groups a day, the number of participants in each group was eight, five, and seven.

There were eight sessions in total. The anagram-embedded control texts were always read in the first three sessions. In the control sessions students stopped reading four times to unscramble anagrams. The control condition's purpose was to serve as a lexical (not metacognitive) task, and to equate both conditions for time on task. Otherwise, the strategies would just have made the experimental condition last longer. The anagram words were comprised of the more difficult words from the section of text that had just been read. The writers chose the ''more difficult'' words by consensus; words had to be longer than four letters and by group consensus considered low frequency for that age group. Readers were always warned before an anagram appeared and prompted – via voiceover and text – if they wished to go back and reread anything (similar to the prompting in the experimental condition with strategies). There were four anagrams in each control text, except in the short practice texts which contained two. The anagram task timed out after 3 minutes and readers were then shown the answers.

The experimental texts contained two verbal and two visual strategies each. In both conditions no one text followed another more than one time. The control texts were written at the same time, and in the same style as the experimental texts. Texts were randomly selected to become either experimental or control. Prior topic knowledge and pre-reading vocabulary scores were not statistically significantly different between the two conditions. Sessions lasted approximately 30 minutes. The entire study lasted 2 weeks.

Trainers. There were two full-time trainers, their job was to troubleshoot problems with the computers and navigation – not to answer content questions. There were three "as-needed" tutors who came in during the one-on-one "What is a strategy" session at time 4. The trainers included the author (who has over 15 years of clinical and research experience teaching reading) and the school's reading specialist. The tutors included two graduate students in either the Educational or Cognitive Psychology programs, and the computer lab manager who was training to become a Principal. All tutors received several sessions of training by the author, and during session 4 everyone followed a protocol booklet (protocol available at www.3D-Readers.com under T3). In designing the protocol booklet to aid with instructor fidelity, much time was spent developing explanations using age-appropriate language to describe and rationalize the use of strategies. It is well known that motivation is a substantial component in whether poor readers will actually stop and work metacognitive strategies on their own. (We also stressed that these were not just strategies to be used on the computer, but anywhere and anytime that students were reading they should stop and use their "new tools".)

Procedure

The time line was as follows:

Time 1: Introductions were made and navigation discussed. The readers read the first short practice control text called "The Wave" I with two embedded anagrams. This text was shorter than the others with only 740 words. (Note: This text appeared again as the experimental practice text at time 4.) It also contained only four vocabulary choices and six final questions.

Time 2: Read Control text.

Time 3: Read Control text.

Time 4: Human tutors sat with the readers for approximately 10 minutes (tutor to reader ratio 1:1, in three instances 1:2) and worked through the protocol booklet explaining to the students what the strategies were and *why* they were important.[2] The readers then read through the experimental version of "The Wave" II practice text. This time, instead of anagrams, readers worked through two of the metacognitive strategies. The text contained one verbal strategy (create a question) and one visual strategy (build a model). No results from this session were included in the analyses, as rereading would surely favor the experimental condition.

Time 5 through Time 8: Readers worked through the four full experimental texts with four strategies in each text.

Results

The results section reports on the following four variables: (1) comprehension assessed with open-ended answers, (2) comprehension assessed with vocabulary gains, (3) rereading assessed with ScrollBacks, and (4) question generation. All analyses were conducted using SPSS version 10.0. All Alpha levels were .05 and bi-directional.

Comprehension – constructed final answers. Once readers exited the text they could not go back to reread. This means that the final questions (and post-reading vocabulary) had to be answered from memory. Students typed in constructed answers to the eight final questions for each text (six in the practice text "The Wave" I). These answers served as the primary measure for comprehension. They were scored using two methods: human experts and HEMA. The author and a graduate student in Educational Psychology scored the answers. A significant interrater reliability was found, Pearson $r = .92$, $P < .001$. All interrater differences greater than 2% were resolved through discussion. Because one scorer was not blind to condition, another reliability measure was used, the assuredly unbiased scoring of the HEMA algorithm. The average human score significantly correlated with the HEMA scores, $r = .79$, $P < .001$. Thus, the functions were very similar. However, the mean intercepts of the two types of scorers differed by a significant average of 13 points ($t > 4.00$). The original HEMA algorithm was propagated with a large range of answers, many of which were too sophisticated for this 6th and 7th grade sample. HEMA did an excellent job overall of assessing an answer's *relative* value, but the baseline was obviously set too high. (Future iterations will remedy this.)

In order to address the hypothesis of whether embedding strategies aided in comprehension, a paired t-test compared final constructed answers during the control condition with final constructed answers

Table 1. Human and HEMA means, SDs, grand means, effect sizes (coher's *d*) and correlations by session and condition for comprehension assessed with constructed answers.

Scoring type	Control 1	Control 2	Control 3	Grand mean (GM) (1–3)	Experimental 5	Experimental 6	Experimental 7	Experimental 8	Grand mean (GM) (5–8)	Effect size (GM)	*r* (GM) Human/HEMA
Human	46.58	36.84	36.37	40.33	51.17	47.91	44.31	46.61	48.18	.45	Control .81**
	(25.19)	(20.99)	(22.07)	(19.58)	(21.92)	(20.73)	(21.59)	(21.17)	(15.69)		
HEMA	28.51	24.71	28.77	27.25	35.95	34.83	36.34	35.92	35.76	.67	Exper. .76**
	(19.73)	(8.51)	(10.17)	(13.59)	(11.96)	(12.57)	(10.29)	(11.86)	(11.67)		

Note: $^{**}P < .01$.

during the experimental condition. Using the human expert scores, the analysis revealed a significant mean difference of 7.84, $t(19) = 3.14$, $P = .005$; using the HEMA scores, the analysis revealed in a significant mean difference of 8.52, $t(19) = 4.60$, $P < .001$. Table 1 shows the relevant statistics. Results support the hypothesis that participants did significantly better on the texts that encouraged and allowed them to work on metacognitive strategies.

Because of the constraints of the within subjects design a further important question presents itself, does comprehension simply improve with time via a practice effect? Is there a linear progression, such that as readers work through more texts their scores simply continue to improve? Figure 3 demonstrates that in each condition there were actually slightly negative slopes associated with time.

To answer this question inferentially, a within subjects growth curve analysis was conducted to compare the changes in comprehension across the control texts with the changes in comprehension across the experimental texts. We wanted to determine if there was a significant interaction between time and condition. The interaction of time and condition on

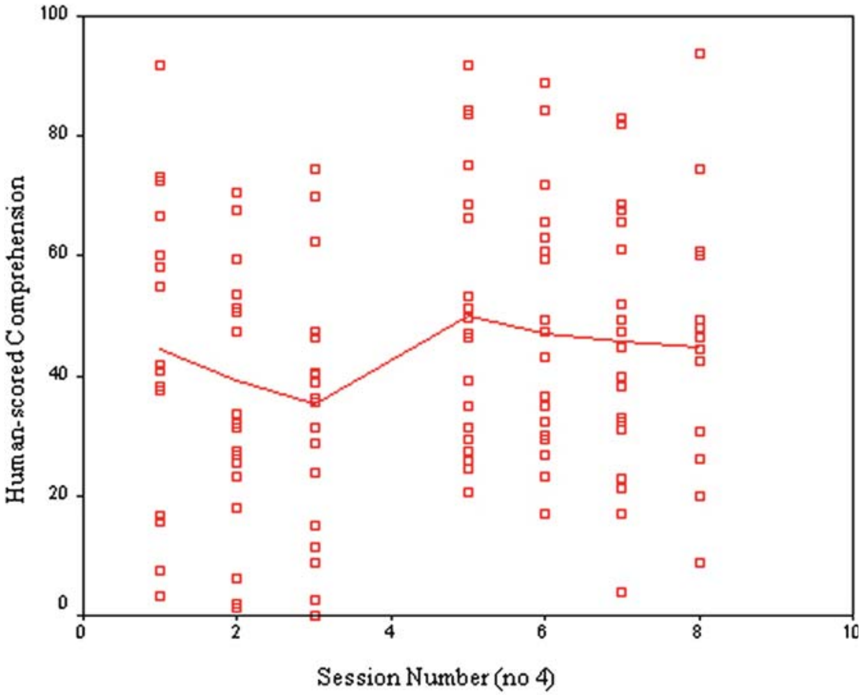

Figure 3. Lowess function fit through the seven sessions: 1–3 are control; 5–8 are experimental.

comprehension was quantified for each student by an individual regression coefficient. A paired t-test was then used to determine if the means of the interaction coefficients differed from zero. The t values were nonsignificant at less than .90. Thus, there is not much evidence for a practice effect that would explain the experimental condition's overall comprehension superiority.[3]

Comprehension – vocabulary gains. The three text authors chose 10 of the most irregular, ambiguous, or age-group-infrequent words from a text. Then the seven words with the most consensual overlap were selected as the vocabulary items. At least five of the seven had to relate specifically to the scientific topic (see Appendix A). Readers answered the seven multiple choice vocabulary items before each text (only four in "The Wave" shorter texts), and they answered the same questions after reading. Table 2 presents the descriptive statistics.

A paired t-test revealed that the average vocabulary gain from pre-reading to post-reading in the control condition was significant, mean gain = 16.43, $t(19) = 4.57$, $P < .001$. In addition, the average vocabulary gain in the experimental condition with strategies was also significant, mean gain = 10.79, $t(19) = 4.22$, $P < .001$. The interaction between gain and condition, i.e., the difference in post-reading minus pre-reading scores between the two conditions, was not significantly different, $t(19) = 1.48$, NS.

We had predicted the experimental group would demonstrate greater gains in the post-reading condition, but this was not the case. Two reasons may explain the results. The first reason may be that the anagram task encouraged readers to focus on the lexical and orthographic levels of the text. Anagrams (and the vocabulary words) were typically the harder words from the text. Thus, extra attention may have been given to the new, more difficult, and often irregular words during reading in order to unscramble them later. This attention could have aided post-reading vocabulary definition, as the vocabulary words were also the more difficult, and irregular words from the text. The second reason may be that the control condition included the short practice text ("The Wave" I) which had only four vocabulary choices and contained somewhat less sophisticated decoys than the full texts. Many of the greatest gains in

Table 2. Vocabulary: percent correct, means, SD, t-tests, and effect sizes.

Conditions	Pre-reading	Post-reading	t-test (pre to post)	Effect size
Control (1–3)	56.39 (16.47)	72.73 (14.93)	$t(19) = 4.57$, $P < .001$	1.04
Experimental (5–8)	60.22 (17.39)	71.01 (17.08)	$t(19) = 4.22$, $P < .001$	.63

vocabulary were seen in this text. The median and mode for post-reading vocabulary were 100% (scored by 13 out of 20 participants) on text 1. On all the other vocabulary pre-reading and post-reading tests the median was never greater than 71%, and no more than four participants ever scored 100%. (Note that the second reading of the Wave practice text II (time 4) was not included in the analyses.)

Finally, how does vocabulary correlate with constructed answers? Table 3 illustrates the significant correlations between vocabulary and constructed final answers as scored by both human experts and HEMA.

Rereading as assessed with ScrollBacks. ScrollBacks were computerized tallies of elective rereading. The majority of ScrollBacks were 0 and 1 (no scrolling to reread, or scrolling up only one line). Because some readers did utilize the double arrow (which scrolled up a block of five lines of text), we had several outlier high scores, ranging from 34 to 64. The distribution of the raw data was non-normal; however, trimming the scores seemed too arbitrary. In order to run parametric tests on these data the numeral 1 was added to each score, and then all scores were natural log transformed. Thus a normal distribution was approximated and 0 remained 0:1 became .69, and 64 became 4.17. Table 4 lists the conditional means and correlations between ScrollBacks, question generation, post-reading vocabulary and constructed answers.

We predicted there would be significantly more ScrollBacks in the experimental condition, and the analyses revealed this to be the case, paired $t(19)=2.16$, $P=.04$. What we did not predict was that on average the participants who were scrolling back more often would be the relatively poorer comprehenders.

Figure 4 illustrates the relationship between the average number of ScrollBacks in the experimental condition minus the number in the control condition and participants' comprehension as assessed in the control condition (before being altered by strategies). The correlation between the difference in conditional ScrollBacks and participants' earlier comprehension was $-.66$ ($n=20$, $P=.002$).

Table 3. Correlations with between post-reading vocabulary and constructed answers.

Post-reading vocabulary and constructed answers	Control	Experimental
Human-scored	.59**	.81**
HEMA-scored	.47*	.58**

Note: $n=20$, $*P=.037$, $**P<.009$.

Table 4. Means, SDs, and correlations between relevant reading variables.

	Mean and SD		Correlations					
	Cont.	Exp.	Voc. Cont. (post)	Voc. Exp. (post)	Human answer – Cont.	Human answer – Exp.	HEMA answer – Cont.	HEMA answer – Exp.
ScrollBack[a] Log$N$ (0–4.17)	0.96 (0.69)	1.35* (0.95)[b]	0.01[c]	−0.69**	0.21	−0.45*	0.28	−0.12
Question generation – Human (0–100%)		74.28 (13.97)		0.54*		0.63**		0.56*
Question generation – HEMA (0–100%)		64.51 [d] ** (7.07)		0.19		0.33		0.32

[a]ScrollBack is natural log transformed. The numeral 1 was added to each score; without transform, the raw control mean was (SD) = 4.65 (5.75), raw experimental mean was (SD) = 8.07 (10.42).
[b]ScrollBack paired t-test, $t(19) = 2.16$, $P = .04$.
[c]The most relevant vocabulary score, which is post-reading is reported. In addition, the vocabulary control correlation with vocab post-reading minus vocab pre-reading difference was .32 (NS). The vocab experimental correlation with vocab difference was .48 ($P = .03$).
[d]Paired t-test between the human and HEMA question generation scores, $t(19) = 3.91$, $P = .001$.

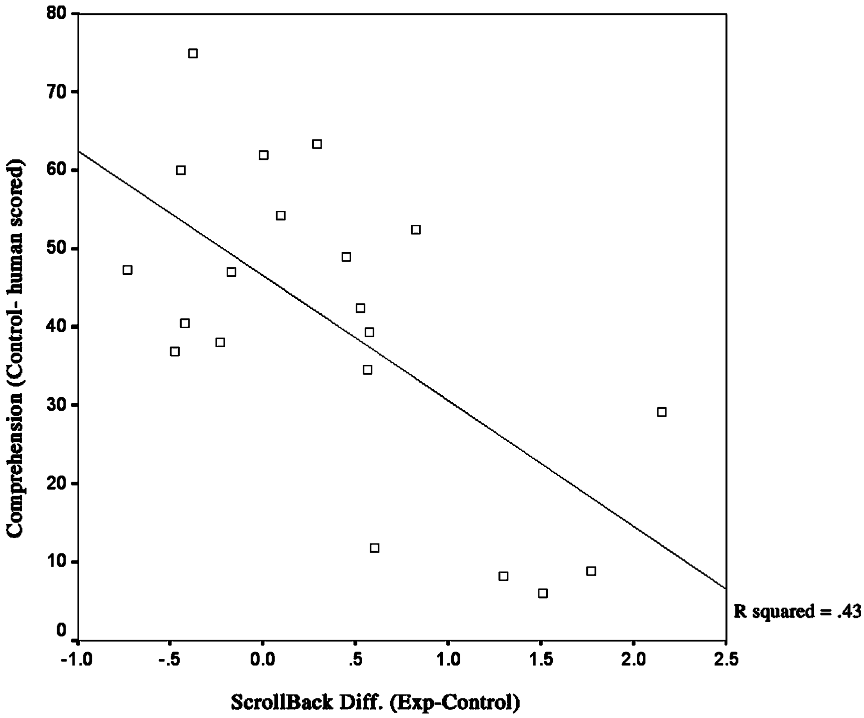

Figure 4. Relationship between ScrollBacks and control condition comprehension.

Question generation. Readers were prompted (via both written text and voiceover) to type in questions during their reading of the experimental texts. They did this once during the experimental practice text (time 4) and twice during the full experimental texts (times 5–8). The readers' self-generated questions were saved in a database and scored by HEMA. However, the scoring of questions is contingent on a more ill-defined problem space than the scoring of constructed answers, and so in this iteration readers did not receive immediate feedback regarding the quality of their questions.

The readers' 165 generated questions were scored three times by (a) the author, (b) a graduate student in Educational Psychology, and (c) the HEMA algorithm. During the one-on-one, time 4 training session readers were instructed that questions that required multi-word answers, and/or began with "how" and "why" were worth more points. A rubric was created for the human scorers which also included these criteria. However, a question that used "how" as a count word, e.g., "How many satellites are in the world?", would not receive as high a score as, "How do signals travel around the world?" The Pearson r correlation between the two human scorers was .69 ($P < .001$). This was not as high as was seen in the scoring of constructed answers. After meeting to discuss discrepancies and come to closer consensus, the scorers' interrater correlation increased significantly to .85 ($P < .000$). Human scorers could not be blind to condition, as question generation was one of the experimental strategies. The Pearson r correlation between the human and the HEMA scores was .61 ($n = 20$, $P = .004$). Again, humans scored the questions at a higher baseline than HEMA did overall. Table 4 shows that the humans scored the questions statistically significantly higher ($M = 74.28$) than HEMA ($M = 64.51$), paired $t(19) = 3.91$, $P = .001$. The algorithm's baseline output will have to be adjusted upwards in this instance as well.

Discussion

The training of reading comprehension using interactive, multi-media Web-based tools is a relatively new field and there are many questions left to be resolved. Hopefully, this study has moved the field forward by addressing two important questions: (1) will training poor comprehenders in visual and verbal strategies result in significant comprehension gains, and (2) will training poor comprehenders in visual and verbal strategies effect elective rereading of texts.

Comprehension. This study demonstrates that scores on the most sensitive comprehension measure – constructed answers to final questions – are significantly higher in the experimental condition with strategies, than in the control condition. This study also presents preliminary evidence that this gain is not simply a linear artifact of time and practice. Although the vocabulary gains were not significantly different between the conditions, the unscrambling of anagrams may have activated some low-level lexical/verbal processes that resulted in vocabulary gains which were not contingent on metacogntive processing. However, it is the use of higher-level verbal strategies and the addition of the visual/imaginal processing that appear to result in an increase in deeper comprehension for the readers, as assessed by the constructed responses. Because the verbal and visual strategies were not separately administered, the question of *which* strategy effects the greatest change cannot be addressed. Many researchers support the use of multi-componential reading programs (NRP, 2000; Pressley, 2000). It may well be the case that the majority of proficient readers use both verbal and visual processes during situation or mental model creation. This is a view supported by Paivio's (1986) Dual Code Theory.

The Dual Code Theory posits three levels of processing or meaning for both the verbal and visual codes. The first level is the *representational* level, this involves the initial activation of one or both code systems. The structure at this level can be "described as the availability in memory of modality-specific logogens and imagens (neuronal structures)" (Sadoski & Paivio, 2001, p. 71). At the second level are *referential* connections which operate between visual and verbal systems. At the third level are *associative* connections which operate within systems connecting imagens and logogens to one another. If comprehension is conceived of as a pattern of neural activation composed of both verbal and visual elements, then a pattern that is both highly activated and veracious represents "good comprehension". The computerized training system in this study may have increased text comprehension for two reasons. First, question generation, by activating the verbal code, may force readers to review current knowledge and ascertain where their knowledge structures are incomplete or fuzzy. Question generation, and answering open-ended questions at the end of the text would certainly activate both representational and associative links in the verbal system. Second, the visual strategy which entails "building a mental model on screen" may activate all three levels of representational, associative and referential links in both verbal and visual systems. Aspects of the text are turned into imagery, lines from the text are repeated, and then readers manipulate and verify where icons should be placed on screen. Activating all three levels and communicating between and within the two verbal and visual systems represents powerful across-the-board cognitive processing. The more practice in effortful cognitive processing that poor comprehenders receive, the more proficient they should become at activating these processes on other texts, and in other literacy situations. Completing these Web-based strategies forces upon the reader self-directed thinking activities. Although this training system supplies a measure of evaluation, the readers themselves must then plan and regulate their cognition, reading, and repair strategies thereafter. Oakhill and Patel (1991) hypothesized that training a visual strategy to their poor comprehenders may have significantly enhanced integration skills (using Dual Code terminology these may be interpreted as both referential and associative connections) or circumvented memory limitations.

ScrollBacks. The analysis of ScrollBacks was edifying. Rereading has been shown to improve comprehension and metacomprehension accuracy in college-age students (Rawson, Dunlosky, & Theide, 2000). On average, the readers in this study used significantly more ScrollBacks in the experimental condition. However, it was the relatively poorer comprehenders who were utilizing the technique more often than the relatively better comprehenders. In the control condition (with the anagrams), ScrollBacks were somewhat positively related to all comprehension variables – vocabulary, human and HEMA scored final constructed answers. However, in the experimental condition ScrollBacks were *negatively* correlated with all the comprehension variables – significantly so with vocabulary and human-scored open-ended answers. Perhaps because the poorer comprehenders came to realize, via the system's feedback, that they were struggling and should avail themselves of the strategy. Perhaps by integrating the system's immediate feedback and their own growing metacognitive awareness, the poorer comprehenders began to more consciously, or at least electively, utilize one of the new tools available to them. Research supports that different readers use text reprocessing strategies differently, e.g., amount of prior knowledge significantly affects text reprocessing (Haenggi & Perfetti, 1992). Interestingly, Haenggi and Perfetti also demonstrate that, on average, college readers show equal benefits in a comparison between three different types of text reprocessing strategies: (1) rereading, (2) rewriting notes, or (3) rereading notes. Taking notes would normally be considered a more effortful strategy and, thus, more beneficial; however, simply rereading the text increased college students' comprehension scores in equal measure.

Importance of findings. The findings are important for several reasons. First, the results demonstrate that Web-based strategy training tools can significantly increase struggling readers' comprehension scores. By allowing readers to create questions as they read, and to manipulate graphics on screen to create stylized simulations of their "mental models", the readers' final comprehension scores increased significantly. The experimental scores increased on average 8% beyond the control scores. When assessing the difference between the control and experimental conditions with human-scored constructed answers, a Cohen's d (or Effect Size – ES) of .45 was found, and in the HEMA-scored condition (with smaller SDs), the ES was .67. As a comparison, in a meta-analysis by Rosenshine and Meister (1994) on the metacognitive intervention Reciprocal Teaching, the median ES with standardized measures was .32.

Second, programming for the visual strategies is expensive and resources dictated the number of media-rich, interactive experimental texts that could be built. Thus, these results are especially unusual and heartening given that the study lasted only two weeks. Duffy et al. (1987) and Meloth (1990) demonstrated that it can take up to 16 weeks for significant higher level metacognitive differences to emerge in poor readers using hard copy materials. Third, this study demonstrates that it is possible to move away from the multiple choice format when testing with computers. There is still refining to be done on the scoring algorithm, but the strong correlations with human scores would lead one to believe that much (though certainly never all) of constructed text can be accurately and reliably scored on-line. This can be said for the larger population of struggling middle school readers that this sample may extrapolate to, it remains to be seen whether the HEMA algorithm will scale up to the average high school and college students who will be using more sophisticated syntax. Fourth, the new generation of formal and informal school assessments expects students to write coherent and orthographically correct short essays. Web-based tools that require constructed text, and that encourage students to practice the extremely valuable skill of writing are important additions to the educational toolbox. Fifth, reading instruction needs more empirical Web-based research. The following quote is from the meta-analysis Report of the National Reading Panel (NRP, 2000): ‘‘Particularly striking in its absence is research on Internet applications as they might be incorporated in reading instruction’’ (p. 6-2).

Future directions. There are two reciprocal directions to focus on in the future: new experiments and further system design enhancements. Controlled studies are needed to ascertain which components benefited text comprehension. One question is how the hybrid text is affecting comprehension. All of the texts address scientific concepts, but these concepts are introduced in an engaging, narrative structure. This narrative structure is more familiar to middle school children than the expository structure. One hypothesis to be tested is whether the hybrid story format is more felicitous for younger readers who may now be struggling with the novel expository format, than it is for older readers (who may find it distracting). Even though the case can be made for multi-componential training packages in the applied domain, it would be of scientific interest to ascertain the ES associated with each individual strategy. In addition, are there interactions between text type (expository, narrative, or hybrid) and metacognitive strategy (verbal, visual, or mixture)?

Future studies will include more participants and texts, and use between-subjects designs to facilitate multiple conditional comparisons without cognitive carryover effects. In addition, a more pedagogically relevant control condition will be included. Instead of unscrambling anagrams, readers will be asked to locate information in the text. This is similar to a control condition used by Lovett et al. (1996). Locating information is a valuable scholastic tool that should not create interference with the question generation or visual strategies. With more participants there would also be enough statistical power to assess for aptitude by treatment interactions. It is hypothesized that practice manipulating images on screen to simulate important textual concepts will especially aid poor visualizers.

The issue of transfer needs to be addressed in a systematic manner. The strategies could be "faded" towards the end of training so that performance could be ascertained on computer texts without embedded strategies. In addition, a far transfer assessment with hard copy texts may be even more ecologically valid.

The computer program will be altered to take into account lessons learned in this study. Some are simple fixes like altering the awkward arrow–click interface. Some fixes are more complex like increasing the baseline on the HEMA algorithm. In future iterations, readers will receive feedback after every submission of a final constructed answer item, instead of after all eight have been submitted. In addition, if the score is less than 90% correct, readers will be automatically taken back to the location in the text where the answer was located and given a second chance. The system will keep track of both answers and a reader who never shows improvement when given a second chance will be flagged for teacher intervention. Further, it would be of interest to track precise placement of ScrollBacks to ascertain what information precedes a ScrollBack. Think-aloud protocols would help to isolate whether readers are engaged in *verification* or *elaboration* processes (Millis & King, 2001).

The new generation of Web-based tools needs to be constructed on top of theory-driven, scientifically, and experimentally sound structures (Reyna et al. 2001). Such tools can be extremely worthwhile for struggling readers who may not be receiving enough of the one-on-one attention and feedback that is so critical to mastering higher level reading comprehension strategies. This study demonstrates that positive change in comprehension and rereading performance can be effected using Web-based training tools.

Acknowledgements

This research was supported by a grant from NICHD R43-HD40696-01. Special thanks are given to Renee Duffy, Matthew Zeidenberg, L. Chris Crawley, Amy Waite, Lori Hillyer, and Gary McClelland.

Notes

1. It should be noted that the effect size dropped to .34 on standardized tests. It is difficult to find comprehension changes with standardized measures (Paris, Cross, &

Lipson, 1984), which is why we did not follow up on this sample's end-of-year standardized measures.

2. It is our goal to eventually create a Web-based system that contains on-line animated tutors. Resources did not allow for the creation of animated tutors in this iteration, human tutors presented the 10-minute strategy introduction in session 4. That 4% of the intervention time was spent with humans does not seriously detract from the claim that this system is primarily "Web-based".
3. Although, it should be noted that the large SD resulted in an effect size of .19, and power was less than .40 to detect a significant difference.

Appendix A

Vocabularies I and II

"These words are going to be in the story. Click on the word or phrase on the right that best matches the word on the left."

1.	Light	a. a form of energy*
		b. not early
		c. pretty
2.	Visible spectrum	a. looks invisible
		b. the colors you can see*
		c. is divisible by three
3.	Reflect	a. small pieces of dirt or other substance
		b. part of a bicycle
		c. to bounce off of a surface*
4.	Astonishment	a. a party
		b. surprise*
		c. a gravel pathway
5.	Absorb	a. take turns
		b. nice
		c. to take something in*
6.	Absence	a. lack of*
		b. sore
		c. a dream-inducing drink
7.	Sample	a. not very clever
		b. an example of something*
		c. a type of light

Appendix B

Shameka Sees the Light

By Julie Fitzpatrick and Mina C. Johnson-Glenberg
Learning Goals: Light travels in waves; absorption and reflection; color is reflected light; color depends on wavelength in source light

"*Take your time and read this story. You will stop four times to work strategies. When you are finished you will be asked eight questions. Use the buttons on the right to move through the text. You can also go back and reread anything you want to.*"

"Shameka! What's up?" asked Jasmine, as she plopped her books down in the dark hallway. Shameka had been waiting by the door for her big sister because she was so excited.

"Look Jas! Mom says if you and I can agree on a color, we can paint our room! Mom and I picked up these cards of sample colors from the paint store this morning. What do you think of this light pink?"

Jasmine took the sample from Shameka's hand and stared thoughtfully at the tiny square of color on the card. "I don't know. . .It looks kind of. . .muddy red, to me," she replied unhappily, as she handed the card back to Shameka. Shameka frowned as she took the card. She walked over to the window where the sun was just setting. She wondered aloud, "That's funny. When I picked it out, I thought it was a nice, pink color. Now, it just looks all red."

"*You are about to work a strategy would you like to reread anything?*"

"*Pretend you are a teacher, type in a question you would ask the class here to see if they were understanding.*"

"Well, bring it in the kitchen and I'll look at it again." Suggested Jasmine, "Man, I could eat a horse. Let's start supper." Shameka called out to her brother loafing in the living room doing nothing, "Michael, turn on the lamp in the living room so Mom can tell we're home when she gets to the corner."

Click. A warm glow lit the other room. The two girls looked at one another in astonishment. Wow, he had actually done their bidding the first time he was asked!

Shameka set the sample card on the counter. Michael came in the kitchen, grabbed the card and ran back into the living room.

"Hey! Come back with that!" Shameka exclaimed.

"Why would anyone pick this color?" he teased, "Pink? Gross!" The two girls turned to look at each other with their heads tilted in puzzlement. Now the sample looked pink again? Michael came back to the doorway. "What? What'd I say?" He asked.

After supper, Shameka and Jasmine discussed the mysterious paint sample with Mom. "To understand what is happening with the paint, you will need to know more about light," said Mom.

"OK," agreed Shameka, "What is light?"

"Light is a form of energy," explained Mom, "Sunlight travels in bursts, called waves. "Um, that's great, Mom, but what does it have to do with the color of the paint sample?" Jasmine was getting impatient.

"We'll get to that," laughed Mom. "The waves act in different ways when they run into something."

"Hold on, do you mean light travels in waves like the kind at the water park?" interrupted Shameka.

Mom replied, "Well, sort of like that. You can still measure the waves from crest to crest, but they are so very teeny tiny that we don't feel pushed around by them like we do by big water waves. The light waves are arranged by how long they are. This is called the visible spectrum. The word visible means we can actually see each wave length as a color. Spectrum tells us they are arranged in some kind of order."

The children were quiet for a while.

"So, what color is made up of the longest waves we can see?" asked Jasmine.

"Red," said Mom, "Then the remaining shorter waves are orange, then yellow, green, blue, indigo, and the shortest waves, that we can still see are violet."

"Hey, that sounds like a rainbow," said Shameka in surprise.

Mom nodded. "Yes, a rainbow shows the visible spectrum, or at least part of it."

"How do you remember all those colors in order, Mom?" wondered Jasmine. She was impressed.

"I just think of the initial letters of the colors as the name Roy G. Biv," said Mom.

"Hey, Mom! That's a great trick! I'm going to use it to remember other important things, like She Has A Man Eating Killer Appetite – **Shameka**!" joked Michael as he threw himself back on the couch.

She gave him one of her famous soon-you'll-be running-for-cover looks, but did not let him distract her from the conversation. "So, are you saying that colors are light, Mom?"

"Yes," her mother replied, "Color is light."

"*You are about to work a strategy would you like to reread anything?*"

Part I – drag color name to appropriate wavelength – assess learning of ROY G BIV

Part II – drag all wavelengths together in box to create white light.

"Now can we get back to how the paint sample changed color?" Jasmine suggested. "OK, do you remember when I said that waves act in different ways when they run into something?" reminded Mom.

"Umm...yeah," answered her three listeners doubtfully.

Mom launched into a careful explanation. "When a light wave hits an object, like a paint sample, or a chair, or anything at all, it can be absorbed by that object, or it can be reflected by that object. What does it mean to be absorbed by something?"

Shameka answered, "You mean like a sponge absorbs water." Mom went on, "That's very close, like a sponge the object can absorb some colors, and the colors that aren't absorbed are bounced off. We say that the light that bounces off the object is reflected. We see the reflected light as the color of the object." Mom smiled at her listeners expectantly.

"*You are about to work a strategy would you like to reread anything?*"

"*Pretend you are a teacher, type in a question you would ask the class here to see if they were understanding*."

They thought about what she had said. Shameka scratched her ear, finally, she asked slowly, ''What if an object absorbs all the light that hits it? What color would it be? No color? Would it be invisible?''

Her mom smiled. ''Remember color is light. What would the absence or lack of light look like?''

''Nighttime!'' shouted Jasmine and Michael at exactly the same time. They laughed at each other.

''True,'' said mom, ''But, we call the night color black. If an object absorbs all light it looks black to us.'' ''Whoa.'' said, Shameka.

''Well, what if the object reflected all the light that hit it? Would it be white?'' asked Michael. All three children looked hopefully at Mom.

''Yes!'' she smiled, ''If the light hitting the object had all the visible spectrum wavelengths in it – if it was white light.''

''OK, OK, suppose you had an object – let's say like that stinky iguana that Mina has. And, he is green. You're saying the iguana absorbs all the wavelengths except green,'' said Jasmine, excitedly. ''That would mean the iguana's only reflecting the green wavelengths, right?''

''*You are about to work a strategy would you like to reread anything?*''

''Yes,'' said Mom, ''that's right. The iguana is only reflecting green and he's absorbing everything else.''

''OK, so if white light hits the iguana he would look green, right?'' continued Jasmine, ''But suppose the light that hit him only contained red wavelengths. Pretend it's nighttime and the traffic light outside the window is stuck on red. NOW what color would the iguana be if only red wavelengths were hitting him?'' Jasmine asked the group while arching her eyebrows.

''Black!'' piped up Michael. ''He would be black because no light would be reflected. No light, no color.'' He looked expectantly at Mom, who nodded.

''Yes! I knew it! '' Michael rejoiced as he did a little victory dance.

Now it was Shameka's turn. ''What if a paint sample absorbed all the wavelengths except mostly pink? Would it be pink in white light and then look more red around sunset when the light seems more red?'' asked Shameka with a confident grin.

''Yes,'' said Mom with satisfaction, ''If there was only a little light coming through the windows at the end of the day, with hardly any of the blue wavelengths, the light would seem redder. The paint sample would look almost red. It would reflect very little pink.''

''Now I get it,'' added Jasmine. ''When Shameka picked the paint sample earlier in the day, the white sunlight had lots of blue, indigo and violet wavelengths in it for the paint sample to reflect. When I looked at it in the setting sun, there were fewer of the shorter blue wavelengths so it only looked totally red.''

''Then I took the sample in here by the light from the lamp, and it looked pink again,'' finished Michael. ''I still say pink is the pits,'' he smirked.

Shameka and Jasmine took another look at the paint sample with Mom. ''Well? What do you think?'' asked Mom.

The girls looked at each other and grinned. ''I think tomorrow we'll be painting!'' shouted Shameka, as she threw the paint sample at Michael, ''and you can help!''

''*Click here to exit the text. You will not be able to go back and reread.*''

Final Open-ended Questions

1. Name the seven main colors that are in the visible spectrum.
2. How does light travel?
3. Are they having horse for dinner? ___Why did Jasmine say that?
4. What does it mean "to loaf"?
5. Why were the girls puzzled when Michael said that pink was gross?
6. Why did the paint sample look different in the evening?

Transfer Questions

7. If a dog reflects ALL the light that hits it in the middle of a sunny day, what color is the dog?
8. Light that is absorbed creates heat in the object that absorbs it. Which would keep you cooler in the summer a white shirt or a black shirt? ____ Why?

References

Amlund, J. T., Kardash, C. A. M., & Kulhavy, R. W. (1986). Repetitive reading and recall of expository text. *Reading Research Quarterly*, *21*, 49–58.

Bell, N. (1986). *Visualizing and verbalizing for language comprehension and thinking* Paso Robles, CA: Academy of Reading Publications.

Bell, N. (1991). Gestalt memory: A critical factor in language comprehension. *Annals of Dyslexia*, *41*, 246–260.

Duffy, G. G., Roehler, L. R., Sivan, E., Rackliffe, G., Book, C., Meloth, M. S., Vavrus, L. G., Wesselman, R., Putnam, J., & Bassiri, D. (1987). Effects of explaining the reasoning associated with using reading strategies. *Reading Research Quarterly*, *22*, 347–368.

FOSS – Full option science program, Lawrence Hall of Science, retrieved October 5:2003, from http://www.lhs.berkeley.edu/foss/.

Gambrell, L. B., & Jaywitz, P. B. (1993). Mental imagery, text illustrations and children's story comprehension and recall. *Reading Research Quarterly*, *28*, 265–273.

Gough, P., & Tunmer, W. E. (1986). Decoding, reading, and reading disability. *Remedial and Special Education*, *7*, 6–10.

Graesser, A., Golding, J. M., & Long, D. L. (1991). Narrative representation and comprehension. In R. Barr, M. L. Kamil, P. Mosenthal, & P. D. Pearson (Eds.), *Handbook of reading research* (Vol. II, pp. 171–205). White Plains, NY: Longman Publishing Group.

Haenggi, D., & Perfetti, C. A. (1992). Individual differences in reprocessing of text. *Journal of Educational Psychology*, *84*(2), 182–192.

Johnson-Glenberg, M. C. (2000). Training reading comprehension in adequate decoders/poor comprehenders: Verbal vs visual strategies. *Journal of Educational Psychology*, *92*, 772–782.

Johnson-Laird, P. N. (1983). *Mental models: Towards a cognitive science of language, inference, and consciousness* Cambridge, MA: Harvard University Press.

Kintsch, W. (1988). The role of knowledge in discourse comprehension: A construction integration model. *Psychological Review*, *95*, 163–182.

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

Levin, J. R. (1973). Inducing comprehension in poor readers: A test of a recent model. *Journal of Educational Psychology*, *65*, 19–24.

Lovett, M. W., Borden, S. L., Warren-Chaplin, P. M., Lacerenza, L., DeLuca, T., & Giovinazzo, R. (1996). Text comprehension training for disabled readers: An evaluation of reciprocal teaching and text analysis training programs. *Brain and Language*, *54*, 447–480.

Mayer, R. E., & Sims, V. K. (1994). For whom is a picture worth a thousand words? Extensions of a dual-coding theory of multimedia learning. *Journal of Educational Psychology*, *86*, 389–401.

Meloth, M. S. (1990). Changes in poor readers' knowledge of cognition and the association of knowledge of cognition with regulation of cognition and reading comprehension. *Journal of Educational Psychology*, *82*, 792–798.

Millis, K., & King, A. (2001). Rereading strategically: The influences of comprehension ability and a prior reading on the memory for expository text. *Reading Psychology*, *22*, 41–65.

Millis, K., Simon, S., & tenBroek, J. (1998). Resource allocation during the rereading of scientific texts. *Memory & Cognition*, *26*, 232–246.

National Center for Education Statistics (NCES). Retrieved August 10, 2002, from www.NCES.ed.gov.

National Reading Panel (Report of the), Teaching Children to Read (2000). Reports of the Subgroups. Retrieved June 5, 2001, from www.nichd.nih.gov/publications/nrp/report.htm.

Oakhill, J., & Patel, S. (1991). Can imagery training help children who have comprehension problems?. *Journal of Research in Reading*, *12*, 106–115.

Paivio, A. (1986). *Mental representation: A dual coding approach* New York: Oxford University Press.

Palinscar, A. S., & Brown, A. L. (1984). Reciprocal teaching of comprehension-fostering and comprehension-monitoring activities. *Cognition and Instruction*, *1*, 117–175.

Paris, S. G., Cross, D. R., & Lipson, M. Y. (1984). Informed strategies for learning: A program to improve children's reading awareness and comprehension. *Journal of Educational Psychology*, *76*, 1239–1252.

Pearson, P. D., & Hamm D. N. (2001). *The assessment of reading comprehension: A review of practices: past, present and future*. Technical Report for RAND Corporation.

Pressley, M. (1977). Imagery and children's learning: Putting the picture in developmental perspective. *Review of Educational Research*, *47*, 585–622.

Pressley, M. (2000). What should comprehension instruction be the instruction of? In M. L. Kamil, P. B. Mosenthal, et al. (Eds.), *Handbook of Reading Research* (Vol. III, pp. 545–561). Mahwah, NJ Lawrence Erlbaum Associates.

Rawson, K. A., Dunlosky, J., & Theide, K. W. (2000). The rereading effect: Metacomprehension accuracy improves across reading trials. *Memory and Cognition*, *28*, 1004–1010.

Reyna, V. F., Brainerd, C. J., Effken, J., Bootzin, R., & Lloyd, F. J. (2001). The psychology of human–computer mismatches. In C. R. Wolfe (Ed.), *Learning and teaching on the world wide web* (pp. 23–41). San Diego, CA: Academic Press.

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

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

Sadoski, M., & Paivio, A. (2001). *Imagery and text: A dual coding theory of reading and writing* Mahwah, NJ: Lawrence Erlbaum Associates.

Stein, N. L., & Glenn, C. G. (1979). An analysis of story comprehension in elementary school children. In R. Freedle (Ed.), *New directions in discourse processing* (pp. 53–120). Norwood, NJ: Ablex.

TerraNova Reading and Language Arts Test (1997), Monterey, CA: CTB/McGraw-Hill.

Tovani, C. (2000). *I read it but I don't get it: Comprehension strategies for adolescent readers* Portland, ME: Stenhouse Publishers.

Trabasso, T., & Magliano, J. P. (1996). How do children understand what they read and what can we do to help them?. In M. Graves, P. Broekvan den & B. Taylor (Eds.), *The first R: A right of all children* (pp. 255–288). New York: Teachers College Press.

Address for correspondence: Mine C. Johnson-Glenberg, Waisman Center, 1500 Highland Avenue, Madison, WI 53705, USA
E-mail: Mina@NeuronFarm.com, johnsonglen@waisman.wisc.edu