Metacognition Learning (2007) 2:89–105
DOI 10.1007/s11409-007-9013-x

SEEK Web tutor: fostering a critical stance while exploring the causes of volcanic eruption

Arthur C. Graesser · Jennifer Wiley · Susan R. Goldman · Tenaha O'Reilly · Moongee Jeon · Bethany McDaniel

Received: 15 December 2006 / Accepted: 29 April 2007 /
Published online: 17 November 2007

Abstract We investigated the impact of a Web tutor on college students' critical stance and learning while exploring Web pages on science. Critical stance is an aspect of self-regulated learning that emphasizes the need to evaluate the truth and relevance of information as the learner engages in systematic inquiry to answer challenging questions. The Web tutor is called SEEK, an acronym for Source, Evidence, Explanation, and Knowledge. The SEEK Tutor was designed to promote a critical stance through several facilities in a computer environment: spoken hints on a mock Google™ search page, on-line ratings on the reliability of particular Web sites, and a structured note-taking facility that prompted them to reflect on the quality of particular Web sites. We conducted two experiments that trained students how to take a critical stance and that tracked their behavior while exploring Web pages on plate tectonics to research the causes of the volcanic eruption of Mt. St. Helens. The SEEK Tutor did improve critical stance, as manifested in essays on the causes of the volcanic eruption, and did yield learning gains for some categories of information (compared with comparison conditions). However, many measures were unaffected by either the presence of the SEEK Tutor or by prior training on critical stance. We anticipate that robust improvements on critical stance and learning will require more training and/or some expert feedback and interactive scaffolding of critical stance in the context of specific examples.

Keywords SEEK · Web · Critical thinking · Tutoring

This work was conducted while Tenaha O'Reilly was a postdoctoral fellow at the University of Memphis.

A. C. Graesser (✉) · M. Jeon · B. McDaniel
Department of Psychology & Institute for Intelligent Systems, University of Memphis,
202 Psychology Building, Memphis, TN 38152-3230, USA
e-mail: a-graesser@memphis.edu

J. Wiley · S. R. Goldman
University of Illinois at Chicago, Chicago, IL, USA

T. O'Reilly
Educational Testing Service, Princeton, NJ, USA

Standards for science education assume that *critical thinking* is an essential component for understanding in science (AAAS 1993). Critical thinking about science requires learners to evaluate the truth and relevance of information, to think about the quality of information sources, to trace the likely implications of evidence and claims, and to ask how the information is linked to the learner's goals and larger conceptual frameworks (Bransford et al. 2000; Halpern 2002; Linn et al. 2004). Critical thinking is needed to achieve deeper levels of learning that involve causal reasoning, integration of the components in complex systems, and logical justifications of claims.

A *critical stance* is an important element of critical thinking that requires the learner to presuppose that the quality of the information is potentially suspect and requires close scrutiny with respect to its truth, relevance, and other dimensions of quality. A critical stance toward scientific information is especially important in the Internet age. The Internet furnishes an endless number of Web pages on almost any topic imaginable, yet there is no control over the quality of the scientific information presented over the Internet. A critical stance is clearly important when students engage in online inquiry tasks and are allowed to use any Internet sites they find as their sources. However, the extent to which readers use a critical stance to guide their own learning and study is not well understood. The present study investigated the process of learning from the Web for the purpose of answering a difficult science question requiring deep causal reasoning—all in the face of a representative mixture of flawed and reliable Web sites.

Cognitive components of self-regulated learning and critical stance for online science inquiry

Successful learning and inquiry at deep levels typically require initiative and decisions on the part of the learner, so metacognition and self-regulation are a critical part of the process. *Metacognition* consists of conscious and deliberate thoughts about one's thoughts, behavior, emotions, and motivation (Flavell 1979; Hacker et al. 1998). *Self-regulated learning* occurs when learners either generate or are assigned specific learning goals and then actively attempt to achieve these goals through a number of phases: constructing a plan, monitoring metacognitive activities, implementing learning strategies, and reflecting on their progress and achievements (Azevedo 2005; Azevedo and Cromley 2004; Pintrich 2000; Winne 2001; Zimmerman 2001). Each of these phases can be decomposed further. For example, metacognitive monitoring can be decomposed into judgments of learning, feeling of knowing, content evaluation, monitoring the adequacy of a strategy, and monitoring progress towards goals. Examples of learning strategies include searching for relevant information in a goal-directed fashion, taking notes, drawing tables or diagrams, re-reading, elaborating the material, making inferences, coordinating information sources, and summarizing content. These various phases are not pursued in a rigid temporal order because there may be dynamic re-planning and because one goal-plan-action episode may be recursively embedded in a superordinate episode. Nevertheless, it is convenient to assume Pintrich's (2000) taxonomy that crosses these phases with the psychological areas of cognition, behavior, motivation/affect, and context.

We view *inquiry learning* as a subclass of self-regulated learning in which question answering is salient as the main goal and shapes the various phases of self-regulated learning. Inquiry learning occurs when learners ask questions (either self-induced or teacher-induced), hunt for answers, evaluate the quality of the answers, and ask related questions in pursuit of answering the main question (Graesser et al. 2005; White and

Frederiksen 1998). The various phases of self regulation are pervasive during inquiry learning. The planning phase is particularly important when the learning goal is ill-structured and the achievement of the goal is not externally evaluated at the time of learning. The metacognitive monitoring phase is critical when learners need to assess what they already know, what they have learned, whether information is relevant to the question, whether they have achieved a sufficiently deep comprehension of a topic, and whether they should re-read or search for further information on the topic (Wiley et al. 2005). Various learning strategies may be employed in response to planning and monitoring phases. For example, readers may engage in re-reading, taking notes, drawing, elaboration with prior knowledge, or attempting to coordinate or reconcile conflicts across information sources.

We view *critical stance* as an aspect of inquiry learning when the truth or quality of the information is called into question. Once again, all phases of self-regulated learning are applicable, with some being more prominent than others. However, the recognition that the reliability of information needs to be questioned adds another level of complexity to how readers might direct their studying behaviors. With this in mind, we have outlined the skills and strategies that we think underlie successful science comprehension and inquiry on the Internet (Goldman et al. 2003; Wiley 2001; Wiley et al. 2007). They include:

(1) Asking deep questions (Graesser and Olde 2003; Graesser and Person 1994).
(2) Gathering relevant information from multiple sources (Britt and Aglinskas 2002).
(3) Evaluating the validity and reliability of information (Brem et al. 2001; Chinn and Brewer 1993).
(4) Integrating information within and across sources (Rouet 2006; Wiley and Voss 1999).
(5) Constructing explanatory inferences to connect consistent information and resolve inconsistencies (Chi et al. 1994; Goldman and Murray 1992; McNamara 2004; Millis and Graesser 1994; Wiley and Myers 2003).
(6) Constructing a coherent causal model of the system (Mayer 2005; Vosniadou and Brewer 1992; VanLehn et al. 2007).

It is widely acknowledged that the proficiency levels for the above skills are very disappointing, even for college students and especially for science. Most students have trouble asking questions (Graesser and Olde 2003; Graesser and Person 1994), calibrating whether they understand the material (Maki 1998), identifying inconsistencies and anomalous information (Otero and Kintsch 1992), strategically searching through hypermedia (Azevedo and Cromley 2004; Hadwin and Winne 2001), generating causal inferences in science (Graesser and Bertus 1998), integrating related information sources (Rouet 2006), and formulating accurate coherent mental models. Brem et al. (2001) have reported that students are generally uncritical of the arguments they read on Web pages. Because these skills are either absent or unspectacular, students need extensive training with expert modeling, scaffolding, and feedback.

The present study investigated whether providing support through a Web tutor can significantly improve critical stance and learning when college students perform an inquiry learning task on science as they explore the Web. We developed a computer Web tutor to scaffold the acquisition of a critical stance to science learning. The Web tutor is called SEEK, an acronym for Source, Evidence, Explanation, and Knowledge. We conducted two experiments that investigated the impact of the SEEK Web tutor on improving college students' critical stance and learning as they searched the Internet for the causes of the volcanic eruption of Mt. St. Helens. Students either were supported in their inquiry learning with the SEEK Web Tutor or were assigned to comparison conditions. During this time, they selected and read Web sites from the list presented on a Google™ search output page. All navigation was confined to the selected sources stored locally, as opposed to their having free access to the Internet. Students were later assessed on several dependent measures of critical stance and learning of the science content. Our expectation was that the SEEK Tutor would lead to improvements on these measures compared with the comparison conditions.

The SEEK Web tutor

The SEEK Tutor was designed to improve college students' critical stance while they searched for information on the Internet in order to complete an inquiry learning task. The learners searched through Web pages on the topic of plate tectonics. Some of the available Web sites were reliable information sources on the topic, whereas others had erroneous accounts of earthquakes and volcanoes. The goal assigned to the college students in the experiments was to search the Web for the purpose of writing an essay on what caused the eruption of Mt. St. Helens volcano. An answer to this question required that they form a deep mental model of the mechanisms of plate tectonics.

Two previous studies motivated the design and testing of the SEEK Tutor in the present study. Both of these were conducted on the same Web pages (covering plate tectonics) and the same main inquiry goal (causes of Mt. St. Helens's volcano eruption). The first study was an exploratory study conducted to better understand how students behave and learn from inquiry tasks using Web pages (Wiley et al. 2007). We found that unsuccessful students exhibited reading behaviour that was similar for reliable or unreliable information. On the other hand, more successful students were better able to discriminate reliable from unreliable information, and their reading behaviour showed a clear preference for reliable information (Wiley et al. 2007). In a second study (Sanchez et al. 2006), we attempted to support students' ability to discriminate reliable from unreliable information. We did this by introducing a training unit on an inquiry learning task on a different topic: to determine whether or not the Atkins Diet is safe. During this 1-h training unit, students received an instruction page which described how one might determine whether a source is reliable and subsequently were required to fill in worksheets for each site. They had to describe each site in terms of its Source, Explanation, Evidence, and how it related to their prior Knowledge. Readers subsequently made reliability ratings about each site; they were also given "expert" rankings to compare with their own judgements. The purposes of this training component were (a) to provide instruction and practice on critical stance as a support for the students when they performed subsequent inquiry tasks using Web pages and (b) to calibrate student understanding of the "reliability" of information. In a second session, a few days after this first task, students were given the Mt. St. Helens inquiry task without prompting or support. Students who had been given the reliability training were more discriminating in their ratings of reliability of the volcano sources, spent more time reading the reliable information about volcanoes, and developed a better understanding of plate tectonics, as compared to students who did not get the training unit beforehand (Sanchez et al. 2006). Therefore, this 1-h training session had a significant impact on both critical thinking and learning.

The SEEK Web Tutor investigated in the present study builds on different aspects of the successful training paradigm of Sanchez et al. (2006). Instead of instruction, practice, and feedback during the training unit on the Atkins diet, the SEEK Web Tutor in the present study prompted students to think about and justify relevant dimensions of critical stance, inquiry, and self-regulated learning during the course of exploring the Web sites on plate tectonics.

This prompting was expected to encourage metacognitive thought, relevant cognitive skills, and learning compared to simply exploring the Web sites without such prompting.

Figure 1 shows the main components of the SEEK Tutor interface. The Tutor fosters students' critical stance with three main facilities. First, there was a *Hint* button on the Google™ search engine page (hereafter referred to as the Google page) which contained suggestions on how to effectively guide students' search. The Google Web page shown in Fig. 1 was a simulated Google page that had titles and URLs for seven Web sites. This Web page was created to follow the layout and interface characteristics of a typical response to a Google query. It was linked to seven sites stored on a local server that were based on real sites found via an actual Google search (see Wiley et al. 2007 for more detail on the learning materials). Three sites were reliable, including sites from NASA (National Aeronautics and Space Administration), PBS (Public Broadcasting Station), and Scientific American. Three were unreliable sites that explained volcanoes and earthquakes by appealing to the stars, the moon, and oil drilling. The seventh site was an ambiguous site that contained reliable information, but had a ".com" URL, so superficial readers might infer that it was unreliable. When the learner clicked on the Hint button, there were spoken messages that gave reminders of the goal of the task (i.e., writing an essay on the causes of the Mt. St. Helens volcano) and suggestions on what to do next (i.e., reading Web sites with reliable information). There were dozens of these messages that the tutoring system randomly selected whenever the Hint button was clicked by the learner. The purpose of this facility was to provide hints that were relevant to the planning phase of self-regulated learning.

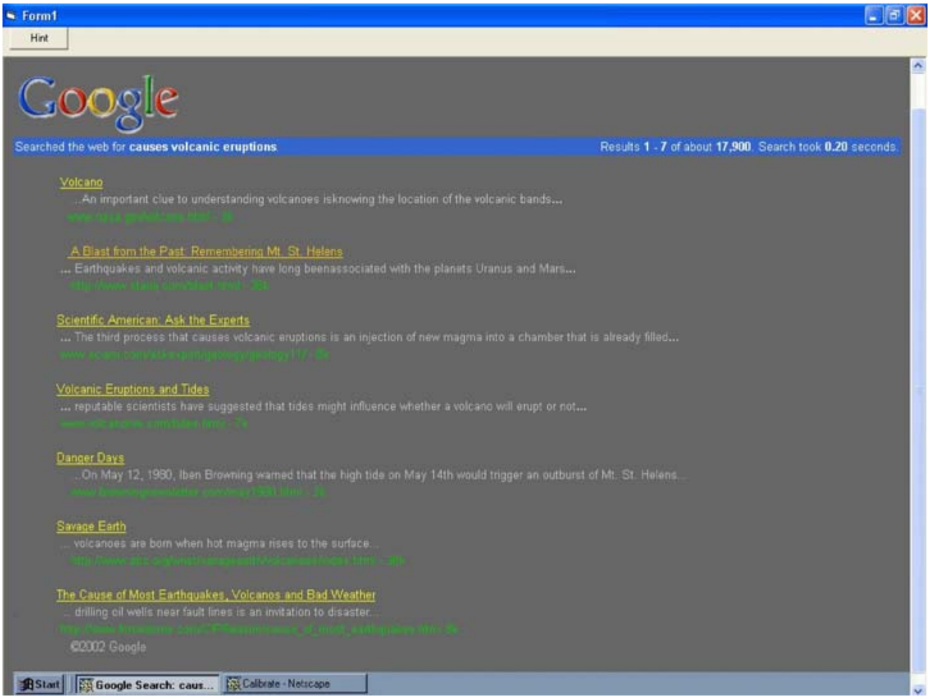

Fig. 1 Interface of mock Google™ page of SEEK Web tutor

The second facility consisted of on-line ratings that were launched by a pop-up window on the Web. The on-line ratings asked students to evaluate the expected reliability of the information in a site by providing a rating and a rationale for their rating. The on-line rating window appeared after the students had first viewed a particular Web site for 20 s. The students were asked to rate the site on reliability using a 6-point scale, where 6 is the most reliable. The participants were also asked to type in a justification for their rating. The purpose of this facility was to encourage the metacognitive monitoring phase in self regulated learning, particularly with respect to evaluating the quality of the Web site when they first encountered the site.

The third facility consisted of a structured note taking interface that had five questions about the reliability of the site that the learner just visited. The Web Tutor launched the pop-up window shown in Fig. 2 when the reader exited a site. These questions were designed to address some of the core aspects of critical stance that were discussed earlier: *Who authored this site? How trustworthy is it? What explanation do they offer for the cause of volcanic eruptions? What support do they offer for this explanation? Is this information useful to you, and if so, how will you use it?* Each question had a Hint button that could be pressed to evoke spoken hints (20 auditory statements per question) to guide the learners on answering each question. The note-taking facility promoted the reflection phase of self-regulated learning by encouraging the learner to think about each of the five core aspects of critical stance and also to articulate verbally the reasons for their ratings.

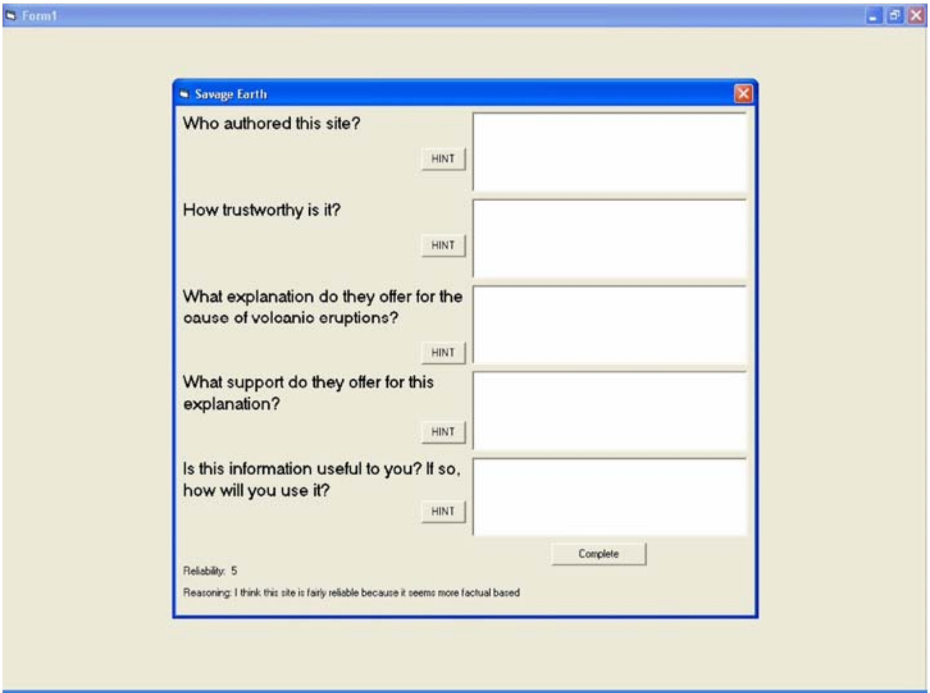

Fig. 2 Structured note-taking facility on SEEK Web tutor

Experiment 1

The participants were randomly assigned to either the *SEEK Tutor* condition or to a *Navigation* comparison condition. We expected that the presence of the SEEK Tutor would be effective in enhancing a critical stance. A critical stance would influence the student's exploration of the Web sites, evaluations of the quality of the Web sites, learning of the content of plate tectonics, and articulation of the causes of the volcano in the essay. To the extent that the SEEK Tutor is effective, there should be better performance in the SEEK Tutor condition than the Navigation condition. On the other hand, it is also conceivable that substantially more training is needed before learners can effectively plan, monitor, and strategically apply a critical stance to science learning, even if they are supported in those behaviours by a Web tutor.

Materials and methods

Participants The sample consisted of 33 psychology undergraduates from the University of Memphis who participated for course credit. The majority of the participants were female Freshman. The participants were randomly assigned to either the SEEK Tutor condition ($n=16$) or to the Navigation condition ($n=17$). The samples did not differ significantly in their science background or Internet usage. This sample size is adequate if we assume a power of 0.90 or higher and an effect size of 1.0, which is the estimated effect size for learning gains from contemporary intelligent tutoring systems (Corbett 2001; Dodds and Fletcher 2004).

Materials and procedure The experiment had three phases: pre-test, online inquiry learning, and post-test. Both the pre-test and post-test assessments were not timed, but the online inquiry task was limited to 50 min. The pre-test consisted of a sentence verification task (true–false) with 30 items. Ten of the 30 statements were true statements based on factual content found on the reliable Web sites. The other 20 statements were false and were broken down into three categories: content (8 items), misconceptions (7 items), and extreme distractors (5 items). The false content statements were incorrect statements altered from correct content on reliable sites (e.g., "New crust is formed where plates converge"). The misconceptions were false statements taken from unreliable Web sites (e.g., "Oil drilling causes volcanic eruptions"). The extreme distractor category consisted of ridiculous statements that were not provided in any of the Web sites (e.g., "Dolphins can predict volcanic eruptions"), but might seem plausible to students with low science literacy.

After completing the pre-test, students were randomly assigned to either the SEEK Tutor or Navigation conditions. Participants in both the SEEK and the Navigation conditions were instructed they would have 50 min to "surf the Web" to gather information to help them write an essay on the causes of the volcanic eruption of Mt. St. Helens. A 5-min warning appeared near the end of the session to inform students of the time they had left.

During the online inquiry task, students were given a mock Google results page with seven pre-selected Web sites (see Fig. 1). Two orderings of the Google results page were created, but both pages contained the same seven sites. Of the 7 sites, 3 sites were reliable sites, 3 were unreliable sites, and 1 was ambiguous. The ambiguous site contained reliable information but the URL (a *.com*) might have suggested the presence of unreliable information to superficial readers. The mean number of Web pages per site was 9.14, ranging from 2 to 19. The pages from these sites were actual pages on the Web, except that we deleted links to any Web pages external to the experimental space. In essence, the Web pages were naturalistic, but we limited the pages to 64 pages of the 7 sites and the pages were stored locally.

Students in the SEEK Tutor condition also filled out the on-line ratings and the structured note-taking facility during the process of studying the Web sites, as discussed in the Introduction. The on-line ratings appeared after the students had viewed a particular Web site for 20 s, whereas the note-taking facility appeared when the students left a Web site to return to the mock Google results page. This facility asked five questions about the Web site (see Fig. 2) that had just been viewed and would not allow the students to move on unless they answered each question. The facility displayed the students' previous note-taking responses at the bottom and contained Hint buttons for each of the five questions. If the students returned to a page, the facility displayed the students' previous responses and they added new information to each of the five response categories.

The post-test phase had multiple assessments. Participants were given 40 min to write an essay that explained the eruption of Mt. St. Helens. The Web pages and the note-taking content were not available to the students while they wrote their essays. Participants were subsequently administered the verification test that had the same 30 statements as the pre-test, but with statements randomized in a different order from that of the pre-test. After finishing the statement verification test, participants ranked the reliability of each of the seven sites on a scale from 1 to 7, where 1 is the most reliable; they also rated information reliability on each site on a six-point scale, where 6 is the most reliable. Finally, the students were given a survey on demographic data, previous science courses, and Internet usage.

Results

Comparisons were made between the Web Tutor condition and the Navigation condition on a number of dependent measures. These measures assessed the process of interacting with the Web sites, critical stance, and assessed learning gains.

Process of studying the Web sites If students explored the Web with a satisfactory critical stance, they should spend more time exploring the reliable sites than the unreliable sites. We tested this by computing a depth measure of how thoroughly the participants explored a particular Web site. Depth was the proportion of Web pages within a site that a participant visited during the Web interaction. We performed an ANOVA on the depth scores as a function of condition (SEEK vs. Navigation) and site status (reliable vs. unreliable). The depth of the reliable sites was significantly higher than unreliable sites, 0.70 versus 0.61, respectively, $F(1, 31)=3.83$, MSE=0.04, $p<0.05$. This result shows that participants did tend to read more deeply into the reliable sites during their reading. However, we also found that the depth scores were lower, not higher, in the SEEK Tutor condition than the Navigation condition, 0.56 versus 0.75, respectively, $F(1, 31)=11.13$, MSE=0.05, $p<0.01$. There was no significant interaction between condition and site status, $F(1,31)=0.003$, MSE=0.04, $p=0.96$. We suspect that fewer Web pages were visited in the SEEK Tutor condition because participants had to spend more time completing the on-line ratings and note-taking facilities, leaving less time to read and study the content. It is therefore not meaningful to compare the total amount of time inspecting the Web sites between the SEEK Tutor and Navigation conditions. The important result is that participants in both conditions allocated more effort to studying the reliable sites than unreliable sites, but this inclination did not differ between the two conditions.

Analysis of ratings and rankings of Web site reliability The ratings and rankings of the Web sites on reliability at the post-test should reflect how well the students acquired a critical stance from the Web interaction. Participants in both conditions were able to significantly differentiate reliable sites from unreliable sites, but there were no significant differences between the two conditions. Mean reliability ratings were significantly higher for the three reliable sites than the three unreliable sites, with means of 4.78 and 3.53, respectively, $F(1, 31)=13.66$, MSE=0.93, $p<0.01$, $\eta^2=0.31$. However, there was no significant main effect of condition $F(1, 31)=0.001$, MSE=1.42, $p=0.97$, nor an interaction between site status and condition, $F(1, 31)=0.63$, MSE=0.93, $p=0.54$. Analogously, mean rankings were significantly lower for the three reliable sites than the three unreliable sites, 3.17 versus 4.74, $F(1, 31)=7.80$, MSE=2.71, $p<0.01$, $\eta^2=0.20$, but there were no effects of condition, $F(1, 31)=0.30$, MSE=0.49, $p=0.59$.

Analysis of pre-test and post-test statement verification to assess learning We computed hit rates for the ten correct statements on the statement verification tests and false alarm rates for the three classes of incorrect statements (i.e., false content, misconceptions, and extreme distractors). A hit rate is defined as p (participant says statement is true | statement is true); false alarm rates are computed as p (participant says statement is true | statement is false). There were no significant differences between conditions (SEEK Tutor vs. Navigation) for any of these measures, but there were significant learning gains. There was a significant increase in hit rates for the ten true content items from pre-test to post-test, with means of 0.76 and 0.88, respectively, $F(1, 31)=15.14$, MSE=0.02, $p<0.01$, $\eta^2=0.33$. However, the learning gains were not significantly affected by the SEEK condition in an ANOVA that crossed condition (SEEK Tutor vs. Navigation) with testing phase (pre-test vs. post-test), $F(1, 31)=1.07$, MSE=0.02, $p=0.31$ and there was no significant interaction between condition and testing phase, $F(1, 31)=0.13$, MSE=0.02, $p=0.72$. When comparing the pre-test and post-test scores, the hit rates increased from 0.74 to 0.86 in the SEEK Tutor condition and 0.77 to 0.90 in the Navigation condition.

One might predict that false alarm rates would decrease from pre-test to post-test if learners take a critical stance and reject unreliable information. However, the results revealed that false alarm rates increased rather than decreased. It appears that exposure to the Web sites increased the learners' ratification of both incorrect information and correct information. Nevertheless, differences did not emerge between the SEEK Tutor and Navigation conditions in any of our analyses of false alarms; there were no main effects of condition and no interaction between condition and test phase. There was a nonsignificant increase in false alarms for the 8 false content items from pre-test to post-test, with means of 0.47 and 0.51, respectively, $F(1, 31)=1.34$, MSE=0.03, $p=0.26$, and for the 5 extreme distractor items, 0.30 and 0.35, respectively, $F(1, 31)=2.48$, MSE=0.02, $p=0.13$. There was a significant increase in false alarms for the 7 misconception items from pre-test to post-test, with means of 0.34 and 0.52, respectively, $F(1, 31)=17.81$, MSE=0.03, $p<0.01$, $\eta^2=0.37$.

Similar analyses were conducted on the d' scores that consider the relationship between hits and false alarms as a metric of how well correct statements can be discriminated from incorrect statements. The d' scores did show increases from pre-test to post-test in 5 out of 6 comparisons when crossing the 2 conditions with the 3 categories of incorrect statements; the overall mean d' scores across the 6 comparisons were 1.16 and 1.51 for the pre-test and post-test, respectively. Therefore, the true–false verification test did show that some learning of the science content did occur during the 50-min session. However, the pattern of means did not significantly favor the SEEK Tutor over the Navigation condition.

Analysis of essays The students' essays on the causes of Mt. St. Helens were first parsed into propositions. These propositions were compared to an ideal template that traced a causal map of volcanic eruptions (see Sanchez et al. 2006; Sanchez and Wiley 2006; Wiley et al. 2007). The causal map template consisted of 15 correct ideas and 6 incorrect ideas (misconceptions) about volcanic eruptions. We computed the proportions of these ideas that were present in the essays. There was no significant difference between the SEEK Tutor and the Navigation conditions for the proportion scores of correct ideas, with means of 0.32 and 0.36, respectively, $F(1, 31)=0.23$, MSE=0.05, $p=0.64$. There was also no significant difference in proportion scores for incorrect concepts, 0.34 and 0.32, respectively, $F(1, 31)=0.04$, MSE= 0.08, $p=0.84$.

The essays were also analyzed from the standpoint of there being expressions of critical stance, that is, propositions relevant to the 5 categories of questions in the structured note-taking facility. The proportion of essays that had a proposition that addressed one or more of the five categories was significantly higher in the SEEK Tutor condition than the Navigation condition, with means of 0.38 and 0.16, respectively, $F(1, 31)=4.23$, MSE= 0.09, $p<0.05$, $\eta^2=0.12$. Therefore, the SEEK Tutor did encourage critical thinking, as reflected in expressions that surfaced in the essays. However, this was the only significant effect that emerged when comparing the SEEK Tutor with the Navigation condition.

Discussion

Our objective was to examine whether 50 min of online inquiry learning with the support of the SEEK Tutor improved undergraduate students' critical stance and learning compared with a search-only Navigation comparison condition. We expected that the SEEK Tutor would encourage college students to change their study patterns and learning as they searched for information on the Internet. We discovered, however, that the presence of the SEEK Tutor did not increase the depth of inspecting reliable Web sites, the ability to differentiate reliable versus unreliable sites, learning gains on a true–false statement verification task, or the quality of essays on the causes of the volcanic eruption. The only encouraging news was that SEEK Tutor did lead to more expressions of critical stance in the essay compared with the Navigation condition. Thus, the SEEK Tutor did cause an increase in talk about critical stance. However, it apparently takes more domain knowledge or more substantial prior training on critical stance before college students are able to improve their ability to evaluate the quality of an information source and to show enhancements in learning.

Experiment 2

Experiment 2 was conducted to test whether providing prior training on critical stance in addition to the SEEK Tutor would strengthen the use of a critical stance among learners. We included a manipulation that attempted to make the SEEK Tutor more effective by providing students with prior training on reliability. We based the instruction on the previous study by Sanchez et al. (2006) in which students were given an example inquiry learning task on the Atkins diet, along with a detailed set of instructions and example Web sites that more thoroughly described and illustrated critical stance. In addition to this written instruction, the Sanchez et al. study made participants complete worksheets and ratings during the Atkins unit, and provided feedback from experts on what their ratings would be after the students provided their own ratings on reliability. The Sanchez et al. study provided approximately 1 h of training on critical stance in the context of the Atkins diet prior to completing the inquiry task on plate tectonics. The enhanced training on critical stance (with the Atkins diet) significantly improved reliability discrimination, reading behaviours, and learning gains on the plate tectonics unit compared with a comparison condition without the enhanced training. Thus, 2 h of the right form of training is sufficient to have a significant impact on outcome measures.

Experiment 2 used an Instruction condition that included only the written instruction about reliability from the Sanchez study. Experiment 2 had a 2×2 between-subjects factorial design with Tutor as one independent variable, and Instruction as the other. For the Tutor variable, participants searched the Web pages either with or without the help of the SEEK Tutor, as in Experiment 1. For the Instruction manipulation, the volcano inquiry task either was or was not preceded by a training unit on the Atkins diet which included some on-line didactic instruction on critical thinking and how to determine the reliability of Internet sites. More specifically, the Instruction conditions included a summary of the critical stance principles included in the structured note-taking facility of the SEEK Tutor, Web pages on the topic of the Atkins low carbohydrate diet, a task of evaluating Web sites, and concrete examples of how the critical stance principles would apply to the diet material. The factorial design allowed us to assess the independent contributions of the on-line tutor and prior instruction on the students' critical stance, study processes, and learning. Perhaps the combination of prior instruction and on-line tutoring would produce effective training of critical stance.

Materials and methods

Participants The participants were 118 college students at the University of Memphis who participated for course credit in an introductory psychology course. The participants were randomly assigned to the four conditions. The majority of the participants were female Freshman and were comparable to the sample in Sanchez et al. (2006). The samples assigned to the four conditions did not significantly differ in gender, science background, or Internet usage. The sample size is adequate if we assume a power of 0.90 or higher and an effect size of 1.0.

Materials and procedure The experiment included four phases: on-line instruction, pre-test, online inquiry learning, and post-test. The pre-test, inquiry task, and post-test phases had exactly the same procedure and materials as in Experiment 1. The difference in the two experiments was the presence or absence of the Instructions as a manipulated variable. The instruction, pre-test, and post-test phases were not timed, but the inquiry learning phase was limited to 50 min.

The Instruction condition was based on the instruction pages from Sanchez et al. (2006). There was a SEEK instruction section that provided didactic instruction on how to evaluate the reliability of Web sites, with content that addressed the five components captured in the note-taking facility of the SEEK Tutor (see Fig. 2). There was an example Web page that described an easy weight loss solution and an example set of answers to the five components. The training component of the Instruction conditions was the same as the Sanchez et al. (2006) study except that participants in the present study did not create worksheets to evaluate a set of 6 Web sites on the Atkin's diet, make ratings, or compare their ratings to experts on diets.

Results

As in Experiment 1, all dependent measures relate to performance on online inquiry learning task about volcanoes. Comparisons were made between the SEEK Tutor and the Navigation condition, between the Instruction and No instruction condition, and the interaction of these manipulations on measures of studying behaviour, evaluation of reliability, and learning.

Process of studying the Web sites As in Experiment 1, we analyzed how deeply the participants examined each Web site. Experiment 2 replicated Experiment 1 in showing lower depth scores in the SEEK Tutor conditions than the Navigation conditions, 0.64 and 0.71, but the effect was less dramatic. The Instruction conditions also had significantly lower depth scores than No-instruction conditions, 0.64 versus 0.70, for reasons that are not understood. The different Tutor conditions and Instruction conditions had these main effects, but there were no significant interactions in these analyses. Analyses were performed on the proportion of time that readers were inspecting the good Web sites as opposed to the bad sites [Good/(Good + Bad)]. This proportion had a mean of 0.62 and did not significantly differ among the four conditions in Experiment 2. These results continue to confirm that readers allocate their study time in a manner that is sensitive to the quality of the Web sites, but these allocations were unaffected by the SEEK Tutor and the Instructions.

Analysis of ratings and rankings of Web site reliability Experiment 2 replicated Experiment 1 in analyses of the participants' perceptions of the reliability of the Web sites. Site reliability ratings and rankings were not affected by the Tutor manipulation or the Instruction manipulation; these independent variables showed no significant main effects or interactions with any variables. Mean reliability ratings were significantly higher for reliable than for unreliable sites, 4.94 versus 3.61; mean rankings were similarly lower for reliable than unreliable sites, 3.23 versus 4.74. We performed follow-up analyses of learners' prior knowledge by segregating those with high versus low pre-test scores on the sentence verification task (using a median split). However, prior knowledge had no significant main effect or interaction in any of the analyses. Once again, it appears that substantially more training on critical stance is needed before college students end up improving their evaluations of site reliability.

Analysis of pre-test and post-test statement verification to assess learning Table 1 presents mean hit rates, false alarm rates and d' scores. The d' score is the purest measure of truth discrimination because it segregates response bias from true discrimination between true and false items. A mixed ANOVA was conducted on the d' scores with Tutor condition and Instruction condition as between-subject factors, and test phase (pre-test vs. post-test) and distractor statement category (false conceptual, misconception, vs. extreme) as repeated measures variables. There was a significant effect of test phase, $F(1, 114)=8.39$, MSE= 1.74, $p<0.05$, $\eta^2=0.07$, which replicates Experiment 1 and confirms that learning gains occurred from pre-test to post-test. There was a significant main effect of distractor

Table 1 Hit rates, false alarm rates, and *d*′ scores for tutor versus navigation conditions in pre- and post-tests of experiment 2

Statement category	Hit rate pre-test	Hit rate post-test	False alarm pre-test	False alarm post-test	*d*′ score pre-test	*d*′ score post-test
10 correct items						
Navigation/instruction	0.73	0.84				
Navigation/no instruction	0.69	0.82				
Tutor/instruction	0.71	0.84				
Tutor/no instruction	0.70	0.79				
8 false content items						
Navigation/instruction			0.52	0.55	0.65	1.03
Navigation/no instruction			0.49	0.53	0.63	1.16
Tutor/instruction			0.54	0.53	0.56	1.17
Tutor/no instruction			0.51	0.54	0.54	0.95
7 misconception items						
Navigation/instruction			0.38	0.54	1.04	1.11
Navigation/no instruction			0.33	0.58	1.19	1.01
Tutor/instruction			0.37	0.58	1.03	1.00
Tutor/no instruction			0.40	0.41	0.86	1.37
5 extreme distractor items						
Navigation/instruction			0.24	0.37	1.55	1.77
Navigation/no instruction			0.24	0.40	1.53	1.61
Tutor/instruction			0.27	0.30	1.47	2.05
Tutor/no instruction			0.24	0.31	1.51	1.77

statement category, $F(2, 228)=76.77$, MSE=0.55, $p<0.05$, $\eta^2=0.40$, a two-way interaction between statement category and test phase, $F(2, 228)=6.64$, MSE=0.34, $p<0.05$, $\eta^2=0.06$, and a four-way interaction, $F(2, 228)=4.06$, MSE=0.34, $p<0.05$, $\eta^2=0.03$. There were no other statistically significant effects in the ANOVA.

Follow-up analyses were performed in the light of the four-way interaction. In the Navigation-with-Instruction condition and the Navigation-with-No-instruction condition, there were learning gains on the false conceptual items from pre-test to post-test, but no learning gains for misconceptions and ridiculous items. For the Tutor-with-No-instruction condition, there were learning gains on the *d*′ scores of false conceptual and on misconception statements, but no effect on extreme distractor statements. For the Tutor-with-Instruction condition, there were significant learning gains for the false conceptual items and extreme distractors, but not misconceptions. Therefore, the participants in the SEEK Tutor condition did do somewhat better on the misconceptions and the extreme distractors but it depended on having the prior training on critical stance in the Instruction condition.

Analysis of essays As in Experiment 1, the participants' essays were parsed into propositions and were compared to an ideal template that reflected a causal map of volcanic eruptions containing 15 correct and 6 incorrect ideas. There were no significant differences between the SEEK Tutor and Navigation when examining proportion scores for correct propositions (0.38 vs. 0.36) and misconceptions (0.20 vs. 0.23). There were no significant effects when examining main effects and interactions with the Instruction manipulation (Instruction vs. No-Instruction).

However, the essays were also analyzed on expressions of critical stance, i.e., whether the essay had propositions relevant to each of the five questions in the note-taking facility (see Fig. 2). A mixed ANOVA was run with Tutor and Instruction as the between-subjects factors and the five critical stance categories as a repeated measures factor. The results revealed that there was an interaction between Tutor condition and critical stance category, $F(4, 111)=2.87$, $p<0.05$, $\eta^2=0.09$, and a marginal main effect for Tutor, $F(1, 114)=3.07$, MSE=0.221, $p=0.08$, $\eta^2=0.03$. Overall, SEEK Tutor participants ($M=0.20$, SD=0.20) mentioned more critical stance content in their essays than did the Navigation participants ($M=0.14$, SD=0.22). Further analyses revealed that one critical stance content category was responsible for the interaction between Tutor and the category. Participants in the SEEK Tutor condition mentioned more terms such as "cause" and "explanation" in their essays than did participants in the Navigation condition, with means of 0.34 and 0.17, $F(1, 114)=9.33$, MSE=0.092, $p<0.05$, $\eta^2=0.08$. The Instructions manipulation had no impact on the critical stance wording in the essays. Therefore, the participants used more causal language in their essays when they had received the SEEK Tutor. The SEEK Tutor did encourage critical thinking, as reflected in such expressions that surfaced in the essays.

Discussion

The SEEK Tutor was purposely designed to support readers in taking a critical stance through the planning, metacognitive monitoring, and reflection phases of self-regulated learning (Azevedo 2005; Pintrich 2000; Zimmerman 2001). The hints on the mock Google pages reminded the learners about their goals and planning, whereas the on-line ratings fortified metacognitive monitoring of the reliability of Web sites, and the structured note-taking facilities facilitated reflection on critical stance components. We had expected there to be improvements in discrimination of reliable and unreliable information, changes in study patterns, and improvements in learning because of these metacognitive supports in the SEEK Web Tutor.

The results from Experiments 1 and 2 support the conclusion that the presence of the SEEK Tutor during an online inquiry learning task had either no effect or only a subtle effect on the various dependent measures. The SEEK Tutor did not improve learners' ability to detect reliable information sources nor their allocation of study resources to more reliable Web sites. These measures are aligned with the metacognitive monitoring phase of self-regulated learning. The college students demonstrated some ability to discriminate reliable from unreliable Web sites and to allocate their study time accordingly, but the SEEK tutor made no dent in improving these skills.

The SEEK tutor did show a small tendency to increase learning for some types of information. For example, the statement verification results showed the Tutor did slightly increase the ability to discriminate between true statements and misconceptions and between true statements and extreme distractors. However, we are not confident that these results will replicate because they were subtle effects and depended on the Instruction condition. SEEK Tutor participants also did not write better essays in terms of covering the 15 core propositions and avoiding the 6 core misconceptions.

However, there was one consistent result that provides some encouraging news. SEEK Tutor participants did have more expressions in the essay with language about causal explanations (such as "cause" and "explanation") compared to controls. Moreover, in follow up analyses (not reported above), we found that those participants in the SEEK Tutor condition who expressed more words in the note-taking facility also benefited most in learning how to discriminate far-fetched false statements from correct scientific statements about volcanoes. So perhaps the act of expressing critical stance information has some relation to learning how to discriminate truth from absurd scientific claims.

The Instruction manipulation proved to have virtually no impact on the measures collected in Experiment 2. This suggests that it is not sufficient to simply present didactic information on critical stance and to present an example Web site with a critique on whether it passes muster with respect to critical stance. Such information delivery does not suffice in getting college students to apply critical thinking strategies and it did not influence learning gains. As discussed earlier, however, another version of the critical stance training was very successful in improving both the use of critical stance and learning (Sanchez et al. 2006; Wiley et al. 2007). In addition to the critical stance training used in the present study, the participants in the previous studies completed worksheets that required: (a) responses about the components of critical stance as they applied to 6 Web sites on low carbohydrate diets, (b) rankings on the reliability of the 6 Web sites and justification for those rankings, and (c) comparison of generated rankings and those provided from a panel of experts. When considering all of the available evidence from the present study and the previous studies we have conducted (Sanchez et al. 2006; Wiley et al. 2007), the training of critical stance is most robustly improved by three cognitive activities that go beyond comprehension of didactic lessons and examples: (1) applying the critical stance components to several examples or questions, (2) articulating decisions, reasons, justifications, and claims about critical stance, and (3) analyzing the feedback from experts on correct decisions, reasons, justifications, and claims.

It appears that there would need to be more training and better scaffolding techniques in our SEEK Web Tutor before robust effects will emerge on the application of critical stance to Web learning. The effects of the SEEK Tutor were subtle or nonexistent on multiple measures, except for the articulation of critical stance principles in the essay. Does this mean that very little can be accomplished in 1 or 2 h of on-line training? Does it mean that intensive training of the SEEK Tutor would be needed on multiple topics and problems before students would become deep learners of science who have a more penetrating critical stance? Increasing the quantity of training is likely to help, but there currently are no data that assess improvements as a function of hours or days of training. We do know, fortunately, that 2 h of training in the Sanchez et al. (2006) yielded significant improvements in both critical stance and learning gains. This suggests that our next version of the SEEK Tutor will need to incorporate the pedagogical components that made the Sanchez et al. treatment effective. In essence, enhancements are needed in training quality, quantity, and/or both.

There is another conclusion that is abundantly clear from the present set of studies. Without developing and using a critical stance, students will acquire both reliable and unreliable information as they attempt to learn science from the Internet. Critical stance is therefore an essential skill that needs to be added to the literacy repertoire, given the growing popularity of Internet inquiry activities that require self-regulatory learning and metacognitive skills. Future work needs to clarify how this important skill can be instilled and supported.

Acknowledgements This research was supported by the National Science Foundation (REC 0126265, ITR 0325428, REESE 0633918). Any opinions, findings and conclusions or recommendations expressed in this paper are those of the authors and do not necessarily reflect the views of NSF. We would like to thank Brian Haynes, Brandon King, and Kristy Tapp for assisting us in developing the computer software, collecting data, and analyzing the data.

References

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

Azevedo, R. (2005). Computer environments as metacognitive tools for enhancing learning. *Educational Psychologist*, *40*, 193–198.

Azevedo, R., & Cromley, J. G. (2004). Does training on self-regulated learning facilitate students' learning with hypermedia. *Journal of Educational Psychology*, *96*, 523–535.

Bransford, J. D., Brown, A. L., & Cocking, R. R. (Eds.) (2000). *How people learn*. Washington, DC: National Academy Press.

Brem, S. K., Russell, J., & Weems, L. (2001). Science on the Web: Student evaluations of scientific arguments. *Discourse Processes*, *32*, 191–213.

Britt, M. A., & Aglinskas, C. (2002). Improving student's ability to use source information. *Cognition and Instruction*, *20*(40), 485–522.

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

Chinn, C., & Brewer, W. (1993). The role of anomalous data in knowledge acquisition: A theoretical framework and implications for science instruction. *Review of Educational Research*, *63*, 1–49.

Corbett, A. T. (2001). Cognitive computer tutors: Solving the two-sigma problem. In *User modeling: Proceedings of the eighth international conference* (pp. 137–147). Berlin: Springer.

Dodds, P., & Fletcher, J. D. (2004). Opportunities for new "smart" learning environments enabled by next-generation web capabilities. *Journal of Educational Multimedia and Hypermedia*, *13*, 391–404.

Flavell, J. H. (1979). Metacognition and cognitive monitoring: A new area of cognitive developmental inquiry. *American Psychology*, *34*, 906–911.

Goldman, S. R., Duschl, R. A., Ellenbogen, K., Williams, S., & Tzou, C. T. (2003). Science inquiry in a digital age: Possibilities for making thinking visible. In H. van Oostendorp (Ed.) *Cognition in a digital age* (pp. 253–283). Mahwah, NJ: Erlbaum.

Goldman, S. R., & Murray, J. (1992). Knowledge connectors as cohesion devices in text. *Journal of Educational Psychology*, *84*, 504–519.

Graesser, A. C., & Bertus, E. L. (1998). The construction of causal inferences while reading expository texts on science and technology. *Scientific Studies of Reading*, *2*, 247–269.

Graesser, A. C., McNamara, D. S., & VanLehn, K. (2005). Scaffolding deep comprehension strategies through Point&Query, AutoTutor, and iSTART. *Educational Psychologist*, *40*, 225–234.

Graesser, A. C., & Olde, B. A. (2003). How does one know whether a person understands a device? The quality of the questions the person asks when the device breaks down. *Journal of Educational Psychology*, *95*, 524–536.

Graesser, A. C., & Person, N. K. (1994). Question asking during tutoring. *American Educational Research Journal*, *31*, 104–137.

Hacker, D. J., Dunlosky, J., & Graesser, A. C. (Eds.) (1998). *Metacognition in educational theory and practice*. Mahwah, NJ: Erlbaum.

Hadwin, A., & Winne, P. (2001). CoNoteS2: A software tool for promoting self-regulation. *Educational Research and Evaluation*, *7*, 313–334.

Halpern, D. F. (2002). *An introduction to critical thinking* (4th ed.). Mahwah, NJ: Erlbaum.

Linn, M. C., Davis, E. A., & Bell, P. (Eds.) (2004). *Internet environments for science education*. Mahwah, NJ: Erlbaum.

Maki, R. H. (1998). Test predictions over text material. In D. J. Hacker, J. Dunlosky, & A. C. Graesser (Eds.) *Metacognition in educational theory and practice* (pp. 117–144). Mahwah, NJ: Erlbaum.

Mayer, R. E. (2005). *Multimedia learning*. Cambridge, MA: Cambridge University Press.

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

Millis, K. K., & Graesser, A. C. (1994). The time-course of constructing knowledge-based inferences for scientific texts. *Journal of Memory & Language*, *33*(5), 583–599.

Otero, J., & Kintsch, W. (1992). Failures to detect contradictions in a text: What readers believe versus what they read. *Psychological Science*, *3*(4), 229–235.

Pintrich, P. R. (2000). The role of goal orientation in self-regulated learning. In M. Boekaerts, P. Pintrich, & M. Zeidner (Eds.) *Handbook of self-regulation* (pp. 452–502). New York: Academic.

Rouet, J.-F. (2006). *The skills of document use: From text comprehension to web-based learning*. Mahwah, NJ: Erlbaum.

Sanchez, C., & Wiley, J. (2006). Effects of working memory capacity on learning from illustrated text. *Memory & Cognition*, *34*, 344–355.

Sanchez, C. A., Wiley, J., & Goldman, S. R. (2006). Teaching students to evaluate source reliability during internet research tasks. In *Proceedings of the 7th international conference of the learning sciences* (pp. 662–666). Bloomington, IN: ACM Digital Library.

VanLehn, K., Graesser, A. C., Jackson, G. T., Jordan, P., Olney, A., & Rose, C. P. (2007). When are tutorial dialogues more effective than reading? *Cognitive Science*, *31*, 3–62.

Vosniadou, S., & Brewer, W. F. (1992). Mental models of the earth: a study of conceptual change in childhood. *Cognitive Psychology*, *24*, 535–585.

White, B. Y., & Frederiksen, J. R. (1998). Inquiry, modeling, and metacognition: Making science accessible to all students. *Cognition & Instruction*, *16*(1), 3–118.

Wiley, J. (2001). Supporting understanding through task and browser design. In *Proceedings of the 23rd Annual Conference of the Cognitive Science Society* (pp. 1136–1143). Hillsdale, NJ: Erlbaum.

Wiley, J., Goldman, S. R., Graesser, A. C., Sanchez, C. A., Ash, I. K., & Hemmerich, J. (2007). Learning science from internet inquiry tasks: The importance of teaching students to discriminate and use reliable sources. Unpublished manuscript, University of Illinois at Chicago.

Wiley, J., Griffin, T. D., & Thiede, K. W. (2005). Putting the comprehension in metacomprehension. *Journal of General Psychology*, *132*(4), 408–428.

Wiley, J., & Myers, J. L. (2003). Availability and accessibility of information and causal inferences from expository text. *Discourse Processes*, *36*, 109–129.

Wiley, J., & Voss, J. F. (1999). Constructing arguments from multiple sources: Tasks that promote understanding and not just memory for text. *Journal of Educational Psychology*, *91*, 1–11.

Winne, P. H. (2001). Self-regulated learning viewed from models of information processing. In B. Zimmerman, & D. Schunk (Eds.) *Self-regulated learning and academic achievement: Theoretical perspectives* (pp. 153–189). Mahwah, NJ: Erlbaum.

Zimmerman, B. (2001). Theories of self-regulated learning and academic achievement: An overview and analysis. In B. Zimmerman, & D. Schunk (Eds.)*Self-regulated learning and academic achievement: Theoretical perspectives* (pp. 1–37). Mahwah, NJ: Erlbaum.