

Scaffolding Critical Thinking on SEEK

Moongee Jeon, Tenaha O'Reilly, Bethany McDaniel, Arthur C. Graesser
Department of Psychology & Institute for Intelligent Systems
University of Memphis
United States
{mjeon|t.oreilly|btmcdan|a-graesser}@memphis.edu

Susan R. Goldman, Jennifer Wiley
Department of Psychology
University of Illinois at Chicago
United States
{sgoldman|jwiley}@uic.edu

Abstract: This paper investigates the impact of a tutor called SEEK on college students' critical thinking about scientific information as they study information on the Internet. Participants searched a set of websites on the causes of the eruption of Mount St. Helens. SEEK was designed to encourage students' critical thinking through several devices: spoken hints on a Google search page, pop-up questions and ratings on dimensions of critical stance (PUR), and a pop-up journal on critical stance (PUJ). We assessed whether SEEK improved students' critical thinking compared with a "search only" (Navigation) control condition. SEEK did improve the critical thinking, as manifested in essays on causes of the volcanic eruption, but did not improve learning gains and students' ability to differentiate the quality of sites. We believe that improving the latter skills requires much more training and some expert modeling of effective inquiry.

Introduction

Current standards for science education assume that critical thinking, or critical stance, is an important ingredient for understanding in science (AAAS, 1993). Critical thinking requires learners to evaluate information actively, to think about the quality of information sources and evidence, to trace the likely implications of evidence and claims, and to ask how the information is linked to larger conceptual frameworks. Readers with a sophisticated level of critical thinking detect low-quality information as well as inconsistencies and contradictions that place them in cognitive disequilibrium (Graesser & Olde, 2003). In order to identify and rectify these flaws in information quality, learners engage in question asking, question answering, and explanatory reasoning that reflect deeper understanding. The inquiry and explanation processes renew equilibrium (Graesser & Olde, 2003) and produce an improved mental model of the subject matter (Graesser et al., 1996, Coté et al., 1998).

A critical stance toward scientific information is especially important in an age of web page proliferation. The internet affords an endless number of web pages on almost any topic imaginable. Unfortunately, there is very limited control over the quality of the scientific information presented over the internet (Goldman & Bisanz, 2002). Readers need to evaluate information actively by asking themselves about the biases and reliability of various information sources, and how information from different sources provides coherent explanations of the content. Readers also need to access and utilize important information to build causal bridging inferences about scientific text and to actively interpret incongruities in text (Graesser & Bertus, 1998, Wiley & Myers, 2003; Wiley & Rayner, 2000).

Empirical studies of learning from text and multimedia materials, however, show that students have difficulty understanding scientific, informational, and technical text (Britton, 1994, Goldman, 1997, Graesser, 1981), and in detecting inconsistencies (Glenberg et al., 1982, Otero & Kintsch, 1992). College students also have difficulty constructing inferences that follow from information explicitly presented in technical expository text (Graesser, 1981, Goldman & Murray, 1992, Noordman et al., 1992), in transferring important concepts to new situations, in handling anomalous information (Chinn & Brewer, 1993; Otero & Kintsch, 1992), and in generating predictions (Millis & Graesser, 1994). Available research strongly suggests that learners need substantial training and scaffolding to acquire strategies of critical thinking. Learners are not equipped to perform self-regulated inquiry without strategic scaffolding (Azevedo & Cromley, 2004, Dillon & Gabbard, 1998, Hadwin & Winne, 2001). Readers often need training to

construct sub-goals, to evaluate the quality of useful information actively, and to differentiate the relevance of accessed information while monitoring primary goals.

In view of these findings, we have recently developed a tutor called SEEK, an acronym for Source, Evidence, Explanation, and Knowledge. SEEK was designed to promote students' critical stance toward scientific information as they search for information on the web. The primary goal of present study was to scrutinize the impact of SEEK on improving college students' critical thinking while they seek for information on the web for the causes of eruption of Mt. St. Helens. The performance of SEEK was compared to a "search only" (Navigation) condition without the aid of SEEK.

A Brief Sketch of SEEK

SEEK is a computer program designed to improve college students' critical thinking while they search for information on the Internet. Figure 1 shows the main components of the SEEK interface.

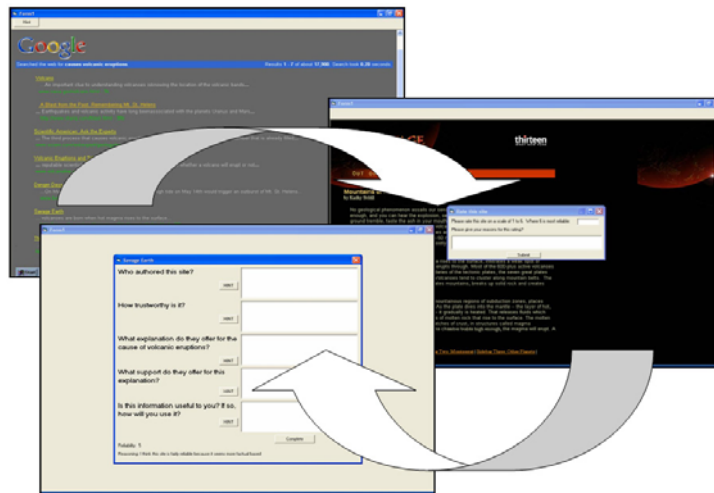


Figure 1: Interface of SEEK

SEEK fosters students' critical thinking with the use of three main devices: (1) A "Hint" button on the Google search engine (upper left screen) which contained suggestions on how to effectively guide students' search, (2) "Pop-up Ratings" (PUR, right screen) which asked students to evaluate a site by providing a rating and a rationale for their rating, and (3) a "Pop-up Journal" (PUJ, bottom screen) which had 5 questions about the reliability of the site students just visited. Example questions are *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? If so, how will you use it?*. Each question had a Hint button which could be pressed to evoke hints (at least 20 auditory statements per question) to guide the learners on answering each question type.

Method

Participants

The sample consisted of 33 psychology undergraduates from the University of Memphis who participated for course credit. The sample included 10 males and 23 females who ranged in ages from 18 to 51 years. The participants were randomly assigned to either the SEEK condition (N=16) or the Navigation condition (N=17).

Materials and Procedure

The experiment had three phases: pre-test, web navigation, and post-test. Both the pre- and post-test assessments were not timed, but the navigation task was limited to 50 minutes. 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 websites. The other 20 statements were false and were broken down into three categories: content (8 items), misconceptions (7 items), and ridiculous distracters (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 seven false statements taken from unreliable websites (e.g., “Oil drilling causes volcanic eruptions”). The ridiculous distracter category consisted of far-fetched statements that were not provided in any of the websites (e.g., “Dolphins can predict volcanic eruptions”).

After completing the pre-test, students in both the SEEK and the Navigation conditions were instructed they would have 50 minutes to “surf the web” to gather information to help them write an essay on the causes of the eruption of Mt. St. Helens. A 5-minute warning appeared at the end of the web navigation to inform students of how much time they had left. Students were given a mock Google results page with seven pre-selected websites. Of the seven sites, three sites were reliable sites, including sites from NASA (National Aeronautics and Space Administration), PBS (Public Broadcasting Station), and Scientific American. Three of them were unreliable sites that contained plausible but incorrect information and explanations on the cause of volcanic eruptions and earthquakes. The seventh site was an ambiguous site containing both reliable and unreliable information. Two orderings of the Google results page were created, but both pages contained the same seven sites.

Students in the SEEK condition were also instructed to fill out the Pop-up ratings (PUR) and the Pop-up journals (PUJ) during the process of studying the web sites. The PUR appeared after the students had viewed a particular website for 20 seconds. The students were asked to rate the site on reliability using a 6-point scale, where 6 is most reliable. The participants were also asked to give a justification for their rating. The PUJ appeared when a student left a website to return to the mock Google results page. The PUJ asked 5 questions about the website that had just been viewed and would not allow the student to move on unless he or she answered each question. The questions regarded source, trustworthiness, the site’s explanation of volcanic eruptions, the site’s supporting evidence, and how relevant the site was to the student. In addition, the PUJ displayed the students’ previous PUJ response at the bottom. Furthermore, the PUJ contained hint buttons for each of the 5 questions. Each button had 20 possible spoken statements that would reframe the PUJ question to help the students formulate their response. Students in the SEEK condition were also given a hint button on the Google page that gave spoken responses reminding them the task was to examine reliable information.

The post-test was a duplicate 30-item statement verification test that was utilized during the pre-test. The statements in the true/false test were randomized in a different order from that of the pre-test. After finishing the post-test, participants were required to write an argumentative essay on the causes of the eruption of Mt. St. Helens for 40 minutes. Participants, then, answered three short questions about plate tectonics, but this task will not be discussed further in this study. Next, the participants rank ordered how reliable each of the seven sites was on a scale from 1 to 7, where 1 is most reliable; they also rated information reliability on each site on a six-point scale, where 6 is most reliable. Finally, the students were given a survey on demographic data, previous science courses, and Internet usage.

Results

Analysis of Pre-Test and Post-Test Scores to Assess Learning Gains

We first computed the proportion of correct responses on the pre-test and the post-test items, including all 30 items. The pre-test scores were not significantly different for the SEEK and the Navigation conditions, with means of .67 and .66, respectively. The post-test scores were also not significantly different, .61 versus .67; if anything, those scores decreased from pre-test to post-test in the SEEK condition. The fact that the overall scores decreased from .66 to .64 from pre-test to post-test suggests that no learning of plate tectonics occurred after exploring the web sites for 50 minutes, at least according to our statement verification test. However, the problem with this conclusion is that overall proportion correct scores are flawed measures because they confound truth discrimination with response bias. It is necessary to conduct signal detection analyses in order to get an adequate assessment of learning gains.

We performed follow-up analyses that segregated the four classes of statements (true content, false content, misconceptions, and ridiculous distracters). These results are presented in Table 1. There were no significant differences between conditions (SEEK versus Navigation) for any of the measures, but there was significant learning gains. There was a significant increase in hit rates for the 10 true content items from pre-test to post-test, with means of .76 and .88, respectively, $F(1, 31) = 15.14$, $MSE = .02$, $p < .001$. A hit rate is defined as: p (participant says statement is true | statement is true). However, the learning gains were not significantly affected by the SEEK condition in an ANOVA that crossed condition (SEEK versus Navigation) with testing phase (pre- versus post-test), $F(1, 31) = 1.07$, $MSE = .02$,

and there was no significant interaction between condition and testing phase. False alarm rates were computed for each category of false items, as defined as: p (participant says statement is true | statement is false). There was no significant decrease in false alarms for the 8 false content items from pre-test to post-test, with means of .47 and .51, respectively, $F(1, 31) = 1.34$, $MSE = .03$; there was no main effect of SEEK condition and no interaction. There was a significant increase in false alarms for the 7 misconception items from pre-test to post-test, with means of .34 and .52, respectively, $F(1, 31) = 17.81$, $MSE = .03$, $p < .001$, but no differences between conditions and no interaction. It should be noted that false alarm rates should decrease with learning, not increase. Therefore, exposure to the web sites increased misconceptions in addition to hits on correct information. For the 5 ridiculous distracter items, there was no significant decrease in false alarm rate from pre-test to post-test, with means of .30 and .35, respectively, $F(1, 31) = 2.48$, $MSE = .02$; there was no significant effect of condition and no interaction. Similar analyses were conducted on the d' scores that compared hits and false alarms as a metric of how well correct target items can be discriminated from distracters (see means in Table 1). The d' scores did show increases from pre-test to post-test (in 5 out of 6 rows), as would be expected if learning occurred. However, the pattern of means never favored the SEEK condition and the differences between conditions were not statistically significant.

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						
SEEK condition	.74	.86				
Navigation condition	.77	.90				
8 false content items						
SEEK condition			.46	.53	.82	1.15
Navigation condition			.47	.49	.86	1.43
7 misconception items						
SEEK condition			.34	.60	1.16	1.01
Navigation condition			.35	.45	1.27	1.63
5 ridiculous distracter items						
SEEK condition			.26	.34	1.49	1.92
Navigation condition			.33	.37	1.35	1.90

Table 1: Hit Rates and False Alarm Rates for SEEK versus Navigation Conditions

Analysis of Ratings and Rankings of Web Site Reliability

Participants in both conditions were able to significantly differentiate reliable from unreliable sites, but there were no significant differences between the two conditions. Mean reliability ratings were significantly higher for the 3 reliable sites than the 3 unreliable sites, with means of 4.78 and 3.53, respectively, $F(1, 31) = 13.66$, $MSE = .93$, $p < .01$. However, there was no significant main effect of condition, nor an interaction between site status and condition. Analogously, mean rankings were significantly lower for the 3 reliable sites than the 3 unreliable sites, 3.17 versus 4.74, $F(1, 31) = 7.80$, $MSE = 2.71$, $p < .01$. However, condition did not have a significant interaction with site status.

Analysis of Essays

The students' essays on the causes of Mount St. Helens were first parsed into propositions. These propositions were compared to an ideal template that traced a causal map of volcanic eruptions. The causal map template consisted of 15 correct ideas and 6 incorrect ideas (misconceptions) about volcanic eruptions. When a proposition corresponded to one of the ideas in the causal map template, we coded it with "1"; otherwise we coded it with "0". The proportions of matching correct propositions were then analyzed. There was no significant difference between the SEEK and the Navigation conditions for the proportion of matching 15 correct concepts, with means of .32 and .36, respectively, $F(1, 31) = .23$, $MSE = .05$. A similar analysis was performed on the 6 anticipated misconceptions. There was also no significant difference between the conditions for the proportion of matching incorrect concepts, with means of .34 and .32, respectively, $F(1, 31) = .04$, $MSE = .08$.

The essays were also analyzed from the standpoint of there being expressions of critical stance, i.e., propositions relevant to the 5 categories of questions in the PURs and PUJs. The proportion of essays that had a proposition that

addressed the 5 categories was significantly higher in the SEEK condition than the Navigation condition, with means of .38 and .16, respectively, $F(1, 31) = 4.23$, $MSE = .09$, $p < .05$. Therefore, the SEEK did encourage critical thinking, as reflected in expressions that surfaced in the essays.

Analysis of Depth

Depth is a measure of how deeply the participants drilled down into the guts of a web site. To calculate the depth rate of each website, we computed the proportion of web pages within a site that a participant visited during training. We performed an ANOVA on the depth scores as a function of condition (SEEK versus Navigation) and site status (reliable versus unreliable). The depth of the reliable sites was significantly higher than unreliable sites, .70 versus .61, respectively, $F(1, 31) = 3.83$, $MSE = .04$, $p < .05$ (one-tailed). This result confirms the prediction that participants would drill down deeper on the better sites. However, we also found that the depth scores were lower, not higher, in the SEEK condition than the Navigation condition, .56 versus .75, respectively, $F(1, 31) = 11.13$, $MSE = .05$, $p < .01$. There was no significant interaction between condition and site status. We suspect that fewer web pages were visited in the SEEK condition because participants had to spend more time on the ratings and justifications in the PUR's and PUJ's, leaving less time to read and study the content.

Conclusions

The primary goal of this paper was to examine whether SEEK improves undergraduate students' critical thinking and learning compared with a "search only" (Navigation) control condition. We expected that SEEK would encourage college students' critical thinking as they search for information on the Internet. SEEK was evaluated on learning gains on a true-false statement verification task, essays on the causes of the volcanic eruption, the ability to differentiate reliable versus unreliable sites, and the depth of inspecting each website.

The results did not support the prediction that SEEK would facilitate learning. There were no significant differences between SEEK and the Navigation conditions when we analyzed the essays and the statement verification judgments (hits, false alarms, d' discrimination scores). It should be noted that learning gains were very small and often nonsignificant when we compared pre-test and post-test scores, so it takes much longer than 50 minutes of exploring web sites to acquire deep knowledge of plate tectonics.

The results were mixed when testing the prediction that SEEK would improve the learners' critical stance. The encouraging news is that SEEK did lead to more expressions of critical stance in the essay, compared with the Navigation condition. Thus, SEEK did cause an increase in talk about critical stance. However, there were no differences between conditions in the students' ability to discriminate good from bad web sites. Perhaps it takes more domain knowledge or more study time before a student is able to evaluate the quality of an information source. The students in the SEEK condition also did not dig deeper into the good web sites compared with the Navigation students, perhaps because they were spending so much time filling out the PUR and PUJ facilities.

Our future work is directed at making SEEK more effective by providing students with a set of instructions and examples that more thoroughly describe and illustrate what is involved in critical thinking. If students have more examples and practice in effectively evaluating their inquiry, they presumably will be more likely to utilize the SEEK in a manner that produces optimal critical thinking and consequently more learning gains.

References

- AAAS (American Association for the Advancement of Science) (1993). *Benchmarks for science literacy: Project 2061*. New York: Oxford University Press.
- Azevedo, R., & Cromley, J.G. (2004). Does training on self-regulated learning facilitate students' learning with hypermedia. *Journal of Educational Psychology*, 96, 523-535.
- Britton, B. (1994). Understanding expository text. In M. A. Gernsbacher (Ed.), *Handbook of psycholinguistics* (pp. 641-674). San Diego, CA: Academic Press.

- 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.
- Coté, N., Goldman, S. R., & Saul, E.U. (1998). Students making sense of informational text: Relations between processing and representation. *Discourse Processes*, 25, 1-53.
- Dillon, A., Gabbard, R. (1998). Hypermedia as educational technology: A review of the quantitative research literature on learner comprehension, control, and style. *Review of Educational Research*, 68, 322-349.
- Glenberg, A.M., Wilkinson, A.C., & Epstein, W. (1982). The illusion of knowing: Failure in the self-assessment of comprehension. *Memory & Cognition*, 10, 597-602.
- Goldman, S. R. & Murray, J. (1992). Knowledge connectors as cohesion devices in text. *Journal of Educational Psychology*, 84, 504-519.
- Goldman, S. R. (1997). Learning from text: Reflections on the past and suggestions for the future. *Discourse Processes*, 23, 357-398.
- Goldman, S. R., & Bisanz, G. (2002). Toward a functional analysis of scientific genres: Implications for understanding and learning processes. In J. Otero, J. A. León, & A. C. Graesser (Eds.), *The psychology of science text comprehension*. Mahwah, NJ: Erlbaum.
- Graesser, A.C. (1981). *Prose comprehension beyond the word*. New York: Springer Verlag.
- 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., & 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., Baggett, W., & Williams, K. (1996). Question-driven explanatory reasoning. *Applied Cognitive Psychology*, 10, S17-S32.
- Hadwin, A., & Winne, P. (2001). CoNoteS2: A software tool for promoting self-regulation. *Educational Research and Evaluation*, 7, 313-334.
- 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.
- Noordman, L. G, Vonk, W., & Kempff, H. J. (1992). Causal inferences during the reading of expository texts. *Journal of Memory & Language*, 31(5), 573-590.
- 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.
- Wiley, J. & Myers, J. L. (2003). Availability and accessibility of information and causal inferences from scientific text. *Discourse Processes*, 36(2), 109-129.
- Wiley, J. & Rayner, K. (2000). Effects of titles on the processing of text and lexically ambiguous words. *Memory & Cognition*, 28, 1011-1021.

Acknowledgments

This research was supported by the National Science Foundation (REC 0126265). 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.