Reading Research Quarterly

International Reading Association
www.reading.org

Comprehending and Learning From Internet Sources: Processing Patterns of Better and Poorer Learners

Susan R. Goldman
University of Illinois, Chicago, USA

Jason L.G. Braasch
University of Oslo, Norway

Jennifer Wiley
University of Illinois, USA

Arthur C. Graesser
University of Memphis, Tennessee, USA

Kamila Brodowinska
University of Illinois, USA

ABSTRACT

Readers increasingly attempt to understand and learn from information sources they find on the Internet. Doing so highlights the crucial role that evaluative processes play in selecting and making sense of the information. In a prior study, Wiley et al. (2009, Experiment 1) asked undergraduates to perform a web-based inquiry task about volcanoes using multiple Internet sources. A major finding established a clear link between learning outcomes, source evaluations, and reading behaviors. The present study used think-aloud protocol methodology to better understand the processing that learners engaged in during this task: 10 better learners were contrasted with 11 poorer learners. Results indicate that better learners engaged in more sense-making, self-explanation, and comprehension-monitoring processes on reliable sites as compared with unreliable sites, and did so by a larger margin than did poorer learners. Better learners also engaged in more goal-directed navigation than poorer learners. Case studies of two better and two poorer learners further illustrate how evaluation processes contributed to navigation decisions. Findings suggest that multiple-source comprehension is a dynamic process that involves interplay among sense-making, monitoring, and evaluation processes, all of which promote strategic reading.

Technological developments in the first decade of the 21st century have made it impossible to ignore the changing face of literacy. People turn to the Web to search for answers to questions and to solve problems that arise in academic, personal, interpersonal, and occupational contexts. Trend data indicate that students of all ages are using the Internet increasingly in their daily lives and particularly when they have to gather information for their schoolwork (Jones, 2002; Lenhart, Rainie, & Lewis, 2001). Increased reliance on the Web for information is potentially problematic because anyone can post just about anything to a website. As a result, information consumers need to critically evaluate information sources for their relevance, reliability, and consistency; they need to synthesize and integrate information across sources to

© 2012 International Reading Association

produce a coherent solution to a problem or an answer to an inquiry question (Bråten, Strømsø, & Britt, 2009; Britt & Aglinskas, 2002; Goldman, 2004, 2010; Lawless, Goldman, Gomez, Manning, & Braasch, in press; Leu et al., 2008; Rouet, 2006; Wallace, Kupperman, Krajcik, & Soloway, 2000). Thus, the Internet world has catalyzed research on learning from multiple information sources, whether they are traditional texts or other multimedia forms.

Comprehending and Learning From Multiple Sources

Theories of comprehension and learning from multiple texts (e.g., Perfetti, Rouet, & Britt, 1999) articulate how models of single-text comprehension (e.g., Kintsch, 1988, 1998) need to be expanded to capture the processing of multiple texts. These early efforts built on investigations of experts reading within their disciplines. When experts read articles in their field to advance their own understanding of a phenomenon, they explicitly consider the reliability of the information sources (Chinn & Malhotra, 2002; Leinhardt & Young, 1996; Moje, 2008; Shanahan & Shanahan, 2008; Wineburg, 1991, 1998). They evaluate the evidence with respect to the methodology used to generate it and how the evidence advances claims, contextualizing these processes within their prior knowledge of the existing literature as well as standards of disciplinary practice (Bazerman, 1985; Berkenkotter & Huckin, 1995; Leinhardt & Young, 1996; Shanahan & Shanahan, 2008; Wineburg, 1998; Yore, Bisanz, & Hand, 2003). This is an inherently intertextual process in which multiple sources of information are juxtaposed with one another, portions are evaluated and selected, and information is integrated as part of a process of updating the expert's mental model about the topic (Goldman, 2004; Hartman, 1995; Rouet, Favart, Britt, & Perfetti, 1997).

In particular, the multiple-documents trace theory (Rouet, 2006; Rouet & Britt, 2011) and its predecessor, the documents model (Perfetti et al., 1999), propose two critical additions to single-text comprehension models: the integrated model, a mental representation that captures connections among the individual representations of single texts; and source node representations that link to each text and represent information about it (e.g., author, reason for creating, publication venue). These additions were intended to represent the key needs for multiple-text comprehension, where information about the documents contributes to the selection, evaluation, and integration of information and its representation in the integrated model. Beyond these needs, as part of multiple-document comprehension, students also need to navigate back and forth among texts and make decisions about when and what to read next. In turn, coordinating among these many processes of navigation, selection, evaluation, connection, and monitoring increases the need for self-regulation skills.

Existing research on single-text comprehension allows us to anticipate likely sources of difficulty for learners, as well as potential differences between better and poorer learners, in multiple-text comprehension situations. Specifically, single-text comprehension research has indicated that more successful readers connect ideas within a text with one another and with relevant prior knowledge, explain the ideas and connections, and actively question their own understanding as well as the text as they attempt to make sense of the information and construct coherent representations (Chi, 2000; Chi, De Leeuw, Chiu, & Lavancher, 1994; Coté & Goldman, 1999; Goldman & Saul, 1990; Graesser, Singer, & Trabasso, 1994; van den Broek, Risden, & Husebye-Hartmann, 1995).

Early studies with multiple sources have suggested that these processes will be even more critical when the connections must be created to construct an integrated mental model, as opposed to single-text situations in which readers can default to representing a single textbase (Britt & Aglinskas, 2002; Wiley & Voss, 1999). Making these connections may also be more challenging in multiple-text contexts because single texts frequently contain cues that signal relationships among different parts of the text (Goldman & Rakestraw, 2000). These signals typically do not exist across multiple texts, so readers must infer and construct them. Indeed, previous studies of multiple-source comprehension have suggested a positive relationship between processing devoted to integrating and synthesizing across sources and subsequent performance on measures of learning and comprehension (Britt & Aglinskas, 2002; Strømsø, Bråten, & Samuelstuen, 2003; Wiley & Voss, 1999; Wolfe & Goldman, 2005).

For example, in one longitudinal think-aloud study, Strømsø et al. (2003) showed that students receiving the highest grades at the end of the year in a university law course had increasingly focused their reading strategies on elaborating the current document that they were reading with respect to text-external sources (i.e., they were making more intertext connections over time). With much younger students (12-year-olds) and a much simpler multiple-source situation, Wolfe and Goldman (2005) reported significant positive relationships between (a) readers' propensities to generate self-explanation inferences that connected within and across multiple texts and (b) their subsequent reasoning about the historical event that was the topic of the two texts.

The emerging research base on navigating and selecting among multiple sources has suggested that high school students and college freshmen use relatively unsophisticated approaches. They generally prioritize content overlap between the task topic and the information source, with limited attention devoted to evaluating the reliability or credibility of the information (Braasch et al., 2009; Britt & Aglinskas, 2002; Kuiper, Volman, & Terwel, 2005; Rouet et al., 1997; Walraven, Brand-Gruwel, & Boshuizen, 2009; Wineburg, 1991). However, as with information integration, learners who display greater sensitivity to the reliability of information sources also tend to learn more of the information (Bråten et al., 2009; Britt & Aglinskas, 2002; Rouet et al., 1997; Sanchez, Wiley, & Goldman, 2006; Stadtler & Bromme, 2004).

For example, Bråten et al. (2009) demonstrated a positive relationship between college students' judgments of texts' trustworthiness and their multiple-document comprehension. However, in their study, as in most others that have demonstrated relationships between sensitivity to differences in reliability among documents and comprehension or learning, judgments of source reliability were made after reading. Thus, prior studies did not address how students actually consider or arrive at evaluations of the reliability of particular documents while they are in the process of reading them.

Finally, because multiple-source comprehension requires managing and tracking different sources, monitoring one's own understanding from multiple sources, and making decisions about what to read next and when, successful learning depends on effective self-regulation. Metacognitive monitoring processes are critical for making effective study choices (Thiede, Anderson, & Therriault, 2003; Thiede, Redford, Wiley, & Griffin, 2012) and are likely to be even more important when learning from multiple sources (Azevedo & Witherspoon, 2009; Coiro & Dobler, 2007; Graesser et al., 2007; Griffin, Wiley, & Salas, in press; Stadtler & Bromme, 2007, 2008).

The Present Research

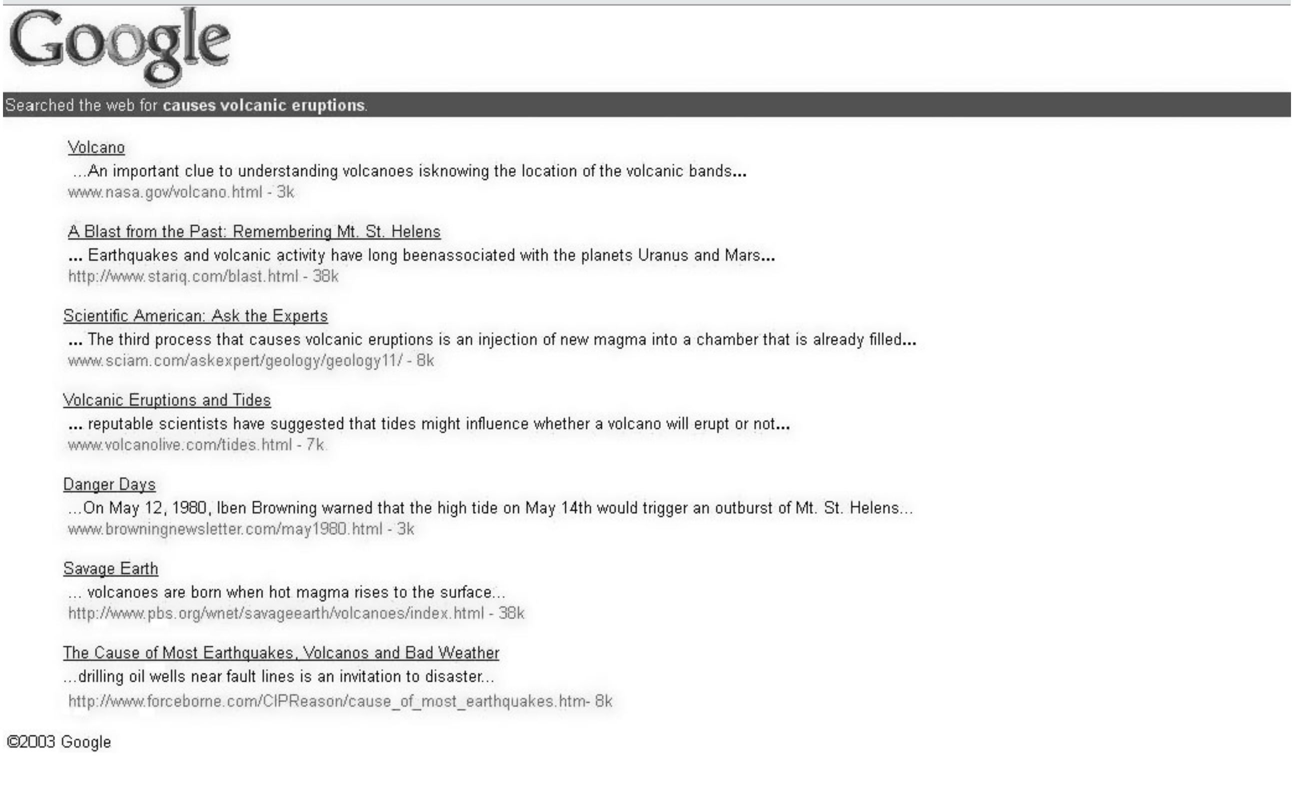
The present study builds on our previously published research on learning from multiple sources of information during a science inquiry task (Wiley et al., 2009). In experiment 1 of Wiley et al.'s study, college students who had relatively low knowledge about the topic that they were asked to research were provided with a set of websites and asked to use the sites to prepare themselves to write an essay explaining the cause of the Mt. St. Helens volcanic eruption that occurred in 1980. The sites were presented in a sheltered web environment via a mock-up of a Google search interface (see Figure). Within the set, the sites varied in their reliability and in the information that they provided about causes of volcanoes. Dependent measures included reading measures derived from navigation logs (duration and number of visits to pages), learning measures (volcano concepts pre- and posttests, number of concepts, and integration in essays), and a source evaluation measure (postreading reliability rankings). (Additional details can be found in the Method section and in Wiley et al., 2009).

The major findings of Wiley et al.'s (2009) experiment 1 were that students who showed more differentiation in their reading behaviors and reliability rankings wrote essays that reflected more correct and integrated causal models of the Mt. St. Helens volcanic eruption. Specifically, those who produced more correct and integrated causal models in their essays, in comparison with those who produced less correct models, spent more time on reliable websites than unreliable ones, returned to pages on reliable sites more than those on unreliable sites, and had larger differences in reliability rankings for reliable and unreliable sites.

Although Wiley et al.'s (2009) data establish important relationships between learning, reading behaviors, and source evaluation, they are limited in that they do not indicate the kinds of processing that participants used to understand the information on each site, how they made decisions about which sites to read, and how they decided whether to continue reading or leave a site. Furthermore, the participants' source evaluation rankings in Wiley et al.'s study were based on measures obtained after participants had finished reading, as well as after writing essays and completing a posttest on volcano concepts. We do not know whether these rankings were based on evaluations made during reading or if they were potentially influenced by the participants having written their essay and responded to a recognition test. However, a subset of the participants in Wiley et al.'s experiment 1 participated in a think-aloud condition. Think-aloud approaches have been used widely in studies of comprehension and self-regulated learning to gain access to how people are processing the materials, interpreting the tasks, setting or revising task goals, and keeping track of their performance (e.g., Chi et al., 1994; Coiro & Dobler, 2007; Coté & Goldman, 1999; Coté, Goldman, & Saul, 1998; Hartman, 1995; Magliano & Millis, 2003).

In this article, we capitalize on the think-aloud protocols to provide a means of addressing sense making and evaluation processes that occur while reading. Based on the prior research outlined earlier, we expected that better learners would spend more time on reliable sites than would poorer learners. We expected better learners to engage in more sense-making processes, such as self-explanation, especially on reliable sites. We further expected that the learner groups might differ in source evaluation and comprehension-monitoring processes. A unique contribution of the analysis of the think-aloud data is in describing the relations that might exist among sense-making, evaluating, and monitoring

Figure. Google Search Results Page That Served as the Interface for the Research Activity



processes during reading and how these may influence online navigation decisions.

Method

Participants

A total of 34 students from the Introductory Psychology Subject Pool at a Midwestern public university participated in the think-aloud methodology condition of the Wiley et al. (2009) study. All met the criteria of low knowledge of plate tectonics and causes of volcanic eruptions, having scored at or below the mean (19 out of 30) on a volcano concepts pretest. As opposed to the contrast of students who wrote the best and worst essays that was described in the Wiley et al. study, in the present study we were interested in exploring the reading behaviors and processes that led to the greatest learning from pretest to posttest. Because of the emphasis in this article on processing related to learning, we used change in performance from pretest to posttest on volcano concepts as the contrast between better and poorer learners.

Better learners were those who had a net increase in correct responses of four or more (range 4–8). Poorer learners were those who showed no net increase or a net decrease in correct responding (range 0–4). Using these criteria, 10 participants (8 male, 2 female) in the think-aloud only sample qualified as better learners and 11 as poorer learners (5 male, 6 female). These samples did not differ in age (better: M = 20.8 years, SD = 4.78; poorer: M = 18.36 years, SD = 0.51; $t(19) = 1.60$, ns), number of science courses taken (better: M = 1.60, SD = 1.65; poorer: M = 0.82, SD = 1.25; $t(19) = 1.23$, ns), number of earth science courses taken (better: M = 0.78, SD = 1.09; poorer: M = 0.45, SD = 0.93; $t(19) < 1$, ns), or rated familiarity (0–100%) with the readings (better: M = 52.22%, SD = 19.22; poorer: M = 50%, SD = 27.20, $t(19) < 1$, ns).

Procedure

Students participated individually in a session of approximately two hours in length. The experiment took place in three phases: the researching/think-aloud phase, the writing phase, and the final assessment phase. For the researching phase, students were instructed to use a set of seven websites that had been returned in a Google search so they could gather information that would allow them to write an argument or a description on "What caused the eruption of Mt. St. Helens volcano?" This essay prompt was manipulated between subjects and is discussed in more detail later. The essay was not actually written until after the end of the researching phase. The seven websites were presented via a Google page mock-up (see Figure). This page acted as the entry point to each of the seven websites and constituted the sheltered web environment in which students conducted their research by selecting and reading the various sites. In addition, the participants were instructed to "think out loud" throughout the researching phase.

Specifically, the students were instructed to read aloud and say what they were focusing on and thinking about. Instructions specified that they should verbalize anything that the information made them think about, including why they decided to read what they were reading and what they thought while reading. Participants were instructed to make comments whenever they wished, but they were encouraged to comment on each sentence they read and each diagram they looked at. As part of the instructions, the experimenter gave example think-aloud comments using an unrelated text on metabolism. We included these examples because our pilot testing indicated that without it participants were unsure of what we meant by thinking aloud and asked whether what they were saying was what we were interested in.

It was problematic to answer this question because regardless of our answer we ran the risk of biasing the types of think-aloud comments they made. Thus, as part of the instructions, the experimenter read through the first paragraph of the metabolism text and modeled thinking aloud with example comments using a prepared script to assure that all participants were exposed to the same examples. The example comments included irrelevant and relevant associations, monitoring statements, predictions, and explanations. They did not include a reference to source quality or credibility. The participant was then asked to "think aloud" after each sentence in a second paragraph. Every participant was given positive feedback for producing comments regardless of the quality or kind of comments that the participant produced. Following this example, the experimenter reiterated the instructions for the Mt. St. Helens inquiry task, and the research phase began. The full think-aloud instruction with examples is provided in Appendix A.

As the participants performed the researching task, the experimenter used a generic think-aloud prompt (e.g., "Can you tell me what you are thinking?") if students were silent for five or more seconds or if they read three sentences without a think-aloud comment. Reading aloud did not count as silence. Prompting was infrequent ($M = 4.48$, $SD = 4.76$ per session).

The researching segment of the experimental session was videotaped and later transcribed. The video recorded the screen and the subject's vocalizations. In this way, we could relate the vocalizations, mouse, and hand gestures to parts of the screen. The printed transcripts of the think-aloud protocols were coded, with reference to the video recordings if there was ambiguity in what the student said or meant.

When participants indicated that they were finished researching (or at the end of the 50-minute time limit), the browser window was closed, and the writing phase began. The students were prompted to write either an argument or a description of what caused the eruption of Mt. St. Helens. They had 40 minutes to write these essays.

The assessment phase followed the writing phase. The essay was saved and closed, and the participants completed a series of tasks, including the same forced-choice volcano concepts test that had been administered at pretest, a reliability-ranking task, and a final survey collecting basic educational and demographic information (age, gender, number of science classes). None of the subjects had trouble finishing the tasks within the time allotted.

Materials and Coding

Causes of Volcanic Eruptions

The topic of volcanic eruptions is particularly well suited to investigating comprehension and evaluation processes in a multiple-source context. There are many interacting causal factors that explain why volcanic eruptions occur. In particular, a complete understanding of the Mt. St. Helens eruption involves integrating scientific information on the physics and chemistry of the earth's crust, the crustal cycle, interpretation of data about earthquake and volcanic occurrences, principles of plate tectonics, and analysis of descriptive information on different types of earthquake and volcanic events (Gobert, 2000; Hemmerich & Wiley, 2002). These concepts and relationships were captured in the causal model shown in Appendix B and used in Wiley et al.'s (2009) study and the present study.

The model was based on prior empirical work on students' conceptions of causes of volcanoes (Gobert, 2000; Hemmerich & Wiley, 2002) and information available from several sources: the U.S. Geological Survey's *This Dynamic Earth: The Story of Plate Tectonics* website (pubs.usgs.gov/publications/text/dynamic.html), a Center for Educational Technologies/NASA Classroom of the Future module on volcanoes (www.cotf.edu/ete/modules/volcanoes/volcano.html), and an introductory geology college textbook (Marshak, 2001). An earth scientist confirmed that the model reflected current scientific understanding of the key concepts and relationships among them. He indicated that although it was possible to depict the model differently, it was certainly a reasonable way to represent the concepts and their interrelationships, especially given the information in the sources.

We used this causal model to guide development of the set of information sources presented to the students, the assessments, and the coding schemes for the various dependent measures. We intentionally adapted the sources to assure that no one source provided the correct or complete answer to the inquiry question, "What caused the eruption of Mt. St. Helens volcano?" (See Wiley, Griffin, & Thiede's, 2005, discussion about critical design features of texts and tests when attempting to assess reading for conceptual understanding.)

Information Sources on Volcanic Eruptions

As indicated above, a mock-up of a Google search results page was the gateway to the seven websites that students could use as information resources for their writing task. The sites were actually stored locally, constituting the sheltered web environment and making it possible to present a finite set of preselected materials that allowed for experimenter control over the content and overlap/uniqueness of information across the sites. At the same time, the mock Google interface preserved the naturalistic feel of an Internet inquiry task. Students were told that the seven sites listed on the Google page were the "top hits" in response to our having entered the search term "causes volcanic eruptions."

The search results page showed the title, the original URL, and a short description of the content of the website for each of the seven sites. The titles were hotlinks to the site, each of which consisted of multiple pages (range 2–8). Most pages contained 300–500 words; included diagrams, maps, or photographs illustrating concepts in the text; and contained sourcing information. Flesch-Kincaid analyses indicated that the verbal text on each of the sites was at the grade 11 or 12 reading level. Sites were listed in two different orders on the Google page, and order was counterbalanced across subjects. Participants tended to read the sites in the order listed, but preliminary analysis indicated that order did not affect any of the measures in this study or in Wiley et al.'s (2009).

The sources themselves were based on actual sites from the Internet. Although the websites looked like the original sources, we made minimal modifications to create discrete pages (needed for the eye-tracking methodology of Wiley et al., 2009; Wiley, Ash, Sanchez, & Jaeger, 2011) and to simplify readability and consistency across sites. Three of the sources were deemed reliable: They are hosted by known, reputable organizations (NASA, *Scientific American*, and the Public Broadcasting Service) and contain accurate and partially overlapping information. The reliable sites contained a total of 5,939 words. As shown in the Figure, the titles of these are *Volcano*, *Scientific American: Ask the Experts*, and *Savage Earth*.

Three other sources were deemed unreliable in that they present incomplete and erroneous accounts of seismic/volcanic activity and are hosted on .com URLs by sources with potentially questionable credibility. Two of these sites came up in a Google search: the astrology site (www.stariq.com) and the fossil fuels site (www.forceborne.com). The astrology site, titled *A Blast From the Past: Remembering Mt. St. Helens*, claims that the location of the planets and stars was responsible for the Mt. St. Helens eruption. The fossil fuels site, titled *The Cause of Most Earthquakes, Volcanos and Bad Weather* (subsequently referred to as *Oil Drilling* in this article), was created by the inventor of an engine that did not run on fossil fuels. The site promotes the engine, and included in the sales pitch is the claim that oil drilling causes volcanoes to erupt. Therefore, we need engines, such as his, that run on alternative fuels. We created the third unreliable website, titled *Danger Days*, from newsletters that had been printed in the 1980s by a corporate forecaster, Iben Browning (www.browningnewsletter.com). Browning claimed that tidal fluctuations allowed him to predict earthquakes and volcanic eruptions. Indeed, he claimed that he had predicted the eruption of Mt. St. Helens.

The formats of the unreliable sites were generally similar to those of the reliable ones. Each provides evidence for their position, but the causal information that the unreliable sites introduce could not be integrated into the model suggested by the reliable sites. A total of five erroneous causes for volcanic eruptions are discussed on the unreliable sites. The unreliable sites contained a total of 5,090 words.

The seventh site is a commercial educational site that had 683 words (www.volcanolive.com) and is titled *Volcanic Eruptions and Tides* (subsequently referred to as *Volcanolive*). The information on this site fits into the causal model supported by the three reliable sites. Although it does not discuss subduction zone volcanoes or mention Mt. St. Helens specifically, it discusses other types of volcanoes. Because of its status as a .com site (rather than a .edu or .org site), people generally have a more difficult time deciding on the reliability of the site. Because of this ambiguity in perception of reliability, this site was not included as either reliable or unreliable in the contrasts testing for differences in processing between reliable and unreliable sources.

Learning Outcome Assessments

The primary measure of learning that was used to define the better versus poorer learners in this study was change in performance from pretest to posttest on a volcano concepts assessment. The assessment was a true-false, forced-choice recognition task that contained 20 items (10 that were true and 10 that were false) that tapped concepts and relations in the causal model of volcanic eruptions. These items represented conclusions that could be generated or refuted by the reading material. An example of an item is "The most violent volcanic eruptions occur in the middle of plates." (This is a false statement that requires the connection of two ideas that were present in the readings.) There were also an additional 10 items, all of which were false, that referred to information that was not part of the causal model. Five of these 10 false items were related to misconceptions that were found in the websites that the students explored during their inquiry (e.g., "The alignment of the stars causes volcanic eruptions"), and five were additional misconceptions not discussed on the sites (e.g., "Sunspots cause volcanic eruptions").

The 30-item assessment, with the items in a different order, was administered as a pretest at the beginning of the semester in a mass testing session. Pretest administration occurred a minimum of one month before the experimental sessions. (For more detail on the creation of recognition tests to assess learning for understanding, see Wiley et al., 2005; Wiley & Voss, 1999.)

Performance on the essay provided three other learning outcome measures: number of core concepts as indicated in the underlying causal model, number of erroneous causes (causes derived from information on unreliable sites), and whether there was any attempt to integrate the concepts of heat and plate movement, or plate movement and pressure buildup. Two independent coders used a detailed scoring rubric to score all the essays. Cohen's Kappa between two raters on individual concepts in the essays was 0.81. A third rater resolved the discrepancies. Two raters agreed 100% on whether the essays reflected conceptual integration.

Reliability Ranking Task

Students were given a paper printout of the Google search page and asked to rank the seven sites based on how reliable they thought they were (1 = most reliable).

Reading Behaviors From the Navigation Logs

During the research phase of the study, we captured moves among pages and time spent on each page, using mouse clicks to indicate moves. We could not capture movement on a page (e.g., scrolling or using the mouse to point to a particular part of the page). From these logs, we honed in on three measures of processing: unique pages visited, returns to pages, and reading time on pages. Averages for these three measures were calculated separately for pages on reliable and unreliable sites.

Processing Measures Derived From the Think-aloud Protocols

The transcripts of the think-aloud protocols included everything that a student verbalized during the research time, including the text read aloud and comments about whatever came to mind while reading. In addition, the angle of the camera captured subjects' gestures toward parts of the screen and mouse movements. We used these to help us annotate referents for deictic pronouns (e.g., "*This* seems to be a picture of what it says *here*") as well as navigation behaviors. The transcriptions were first parsed into comments, defined as the speech burst following the reading of a sentence or group of sentences. This definition of comment corresponds to the coarse grain size used by Chi et al. (1994) in coding verbal protocols. Two raters independently parsed a 10% subset of the protocols with high reliability, achieving a Cohen's Kappa of 0.92. Disagreements were resolved in discussion.

A given think-aloud comment can reflect several different kinds of thoughts about, or processing of, information. To capture these, each comment was parsed into events, where events reflect different kinds of processing or ways to think about the information. Generally speaking, an event corresponds to a clause, sentence, or utterance unit (Chafe, 1994). Our process follows procedures used by Coté et al. (1998) and Wolfe and Goldman (2005) and corresponds to Chi et al.'s (1994) use of the idea unit. Inter-rater agreement on event identification was 92%, with disagreements resolved in discussion.

Following the identification of events, coding for processes proceeded in two passes. In the first pass, eight types of processing events were used in coding the comments: repetition/paraphrase, surface connection, self-explanation, irrelevant association, prediction, monitoring, information/source evaluation, and navigation. The first six categories were based on types of processing events that we identified in our previous work on single texts (Coté et al., 1998; Coté & Goldman, 1999). The first five represent comments about the context of each text and attempts to represent it. Monitoring refers to assessments of one's own understanding. Information/source evaluation was modified from our previous work to include comments about the sites themselves (e.g., author, URL). Navigation was added to reflect statements related to movement among pages and sites. Table 1 provides definitions, and Table 2 provides examples of the various types of comments, how we defined events within comments, and annotations illustrating application of the coding categories to the events.

Two coders independently scored eight protocols parsed into think-aloud comments using the eight categories. Cohen's Kappa was 0.84 for categorizing events. The remaining protocols were each scored by the same two raters, and coding decisions were compared. Any disagreements were resolved in discussion. Comments that were unintelligible were not coded. In addition, within the information/source evaluation and navigation categories, several subcategories were distinguished

Table 1. Descriptions of Think-Aloud Processing Categories

Category	Description
Repetition/paraphrase	Repetition or restatement of the gist of a segment in the text without adding additional information
Self-explanation	Elaboration, interpretation, or reasoning with or about a focal segment; brings new information to the focal segment, including relating it to prior knowledge or information in other information segments
Surface connection	Vague references to previously read information but without any other type of processing (e.g., "I just read that"; "I heard that somewhere before") or that connected part of the text to a diagram on the same website but again added no additional information (e.g., "That's in the diagram"; referring to a concept in the verbal text)
Irrelevant association	Associations to the content of the focal segment that bear little or no obvious relevance to the task: These often relate personal experiences that do not contribute to understanding the information in the context of the inquiry task.
Prediction	Statements about what the learner expects to find out next or what the next segment of text is likely to be about
Monitoring	Statements that confirm comprehension or indicate lack of comprehension, or awareness of prior knowledge (e.g., "I didn't know that")
Information/source evaluation	Judgments about some aspect of the sources, including relevance of the content, consistency with other information, author credentials, credibility, style, or appearance
Navigation	Descriptions of movement within or across pages, including readers' intentions about where to go next, why they wanted to go there or what they were looking for (goals), and reasons for leaving pages that they were in the process of reading

(described in the Results section). Cohen's Kappas for subcategory coding were above 0.86.

In a second pass, we coded for a new category of comments to explore possible differences in the kinds of connections that learners were making across sites: intertext connections. These comments referred to sites other than the one the subject was reading, such as, "This goes directly back to what I read about the oceanic plate going below the continental plate." These comments often overlapped or were integral to several of the first-pass processing categories, especially surface connections but also self-explanations, evaluation, and navigation. To avoid inappropriately inflating the number of processing events, we analyzed these separately from the first-pass coding categories. Cohen's Kappa on a sample of five protocols was 0.82, with disagreements resolved in discussion.

Results

The first section of results provides a preliminary set of quantitative analyses that establishes that the contrasts between better and poorer learners that are the focus of this article replicate the general findings of Wiley et al.'s (2009) experiment 1. In the second section of the results, we explore the processing revealed by the think-aloud analyses. For parametric tests, the dependent measures were submitted to paired samples *t*-tests or ANOVA, in which learner group was the between-participants variable and site reliability a two-level (reliable, unreliable) repeated-measures variable. For all ANOVAs in this article, assumptions of normality and homogeneity of variance were met unless otherwise specified, and significant interactions were followed up by simple effects tests. When appropriate, effect size estimates (Cohen's *d* or η_p^2) were computed to reflect the amount of variance in the sample accounted for by the effect (Cohen, 1988). When assumptions of normality and homogeneity of variance were not met, nonparametric statistics were performed as appropriate (e.g., chi-squares, Wilcoxon signed-rank tests with effect sizes computed as *r*).

Basic Replication of Patterns From the Larger Sample

The data in Table 3 show that patterns of performance for the better and poorer learner groups used in this study were generally consistent with the larger sample in Wiley et al.'s (2009) experiment 1. The first several rows of Table 3 report the learning outcome measures. The first analysis shown in the table simply validates that the learner groups did not differ in their volcano concepts pretest scores. The second analysis indicates that they had significantly different gain scores. The third analysis shows that essays produced by the two learner groups did not differ on inclusion of single correct concepts. However, five of the better learners, as compared with only one of the poorer learners, produced essays that integrated concepts (i.e., included relations between the

Table 2. Examples of Coding of Think-Aloud Comments

Focal Segment Being Read[a]	Think-Aloud Comments Parsed Into Events	Coding Category (Annotation)
"The molten rock or magma collects in weak patches of crust."	(1) "So, it's like when you pour on pizza sauce, and it dips into the crust."	Event 1: Self-explanation (Makes analogy to everyday life)
"If the pressure in the magma chamber builds high enough, the magma will erupt. A volcano is born."	(1) "Which makes complete sense. (2) It's like a zit almost because all the stuff is like inside, and it goes (makes exploding sound). That's really gross."	Event 1: Monitoring (Monitors sensibility of segment in light of prior knowledge) Event 2: Self-explanation (Draws on prior knowledge and makes analogy)
"When a part of the earth's upper mantle or lowered crust melts."	(1) "We just read stuff about this. (2) I don't think I should have to read it anymore. (3) We're gonna go on down to the next paragraph."	Event 1: Surface connection (Notices similarity of this information to something just read) Event 2: Evaluation (Devalues reading the information because it is redundant with something already read) Event 3: Navigation (Narrates move to the next section)
"Everybody is familiar with ocean tides that cause the ocean level to go up and down, usually twice a day."	(1) "Like sort of when the turtles come and lay their eggs with the tide, and then they go back."	Event 1: Irrelevant association (Draws on prior knowledge, but turtles laying eggs does not contribute to understanding tides and ocean levels or volcanoes)
"The 'skin' of Earth, called the crust, is very thin like in the skin of a peach."	(1) "All right, so I think (2) this will tell me a little bit about what will make it erupt."	Event 1: Prediction (Anticipates what next section of text will be about) Event 2: Evaluation (Evaluates content relevance)
"Scientists cannot exactly predict eruptions."	(1) "It's about scientists and predictions, (2) so I'll just scan some more."	Event 1. Paraphrase (States the topic of the sentence) Event 2. Navigation (Decides to scan further in the text)
"Earthquakes and volcanic activity have long been associated with the planets Uranus and Mars, so one might expect to find these two planets prominent."	(1) "But it doesn't have anything to do with the topic, (2) so I'm gonna keep going (giggles)."	Event 1: Evaluation (Decides that content is not relevant) Event 2: Navigation (Decides to move on)
"This force combined with the centrifugal forces could (if we continue to abuse our planet) shake the earth to pieces."	(1) "The wording doesn't make me believe him. 'Shake the earth to pieces' doesn't seem very educated. Ummm. (2) I'm starting to question the author."	Event 1: Evaluation (Evaluates the credibility of the information content) Event 2: Evaluation (Questions the reliability of the author)

[a]The segments in column 1 are taken from the following websites (in order): *Savage Earth* (http://www.pbs.org/wnet/savageearth/volcanoes/index.html); *Savage Earth*; *Scientific American* (www.sciam.com/askexpert/geology/geology11/); *Volcano Eruptions and Tides* (www.volcanolive.com/tides.html); *Volcanoes* (www.nasa.gov/volcano.html); *Scientific American*; and *A Blast From the Past* (http://www.stariq.com/blast.html).

concepts of heat and plate movement, or plate movement and pressure buildup: $\chi^2(1) = 4.3, p = .04$). Furthermore, only one of the better learners, as compared with six of the poorer learners, included at least one erroneous cause in their essays ($\chi^2(1) = 4.67, p = .04$).

Table 3 also indicates that the dependent measures derived from the navigation logs confirm that the present subset of learners mimicked the overall findings of the larger sample. First, it is important to note that no differences were seen between learner groups in the total amount of time spent on the task. In terms of relative time spent on pages of reliable versus unreliable sites, better learners had almost a 4:1 ratio, a significantly larger ratio than the 1.4:1 ratio for the poorer learners. This led to a significant interaction between learner group and site reliability for number of pages visited at least once (unique visits, $F[1, 19] = 11.50, p = .003, \eta_p^2 = 0.38$) and returns to pages ($F[1, 19] = 6.30, p = .02, \eta_p^2 = 0.25$), with better learners having higher means, as compared with poorer learners, on reliable websites but not on unreliable ones. Thus, the better learners were more likely to go to pages on reliable sites and to return to them than poorer learners were.

Likewise, the reliability rankings show that better learners' rankings tended to differentiate between reliable and unreliable sites to a greater degree than those of the poorer learners, although the effect was not statistically significant. Thus, although it was in the same direction as the larger sample, it was not as robust, possibly due to limited power from the modest sample size.

In the larger sample of Wiley et al.'s (2009) study, a small effect was seen for the essay writing manipulation on some measures, with an argument prompt leading to slightly better performance. Yet, even in the larger

Table 3. Learning Outcomes, Reading Patterns, and Differences in Better and Poorer Learners' Reliability Rankings of Websites

Measure	Better learners M (SD)	Poorer learners M (SD)	Statistical test	Effect size
Learning outcomes				
Volcano concepts pretest	17.20 (2.25)	18.0 (1.26)	$t(19) = −1.02$	
Learning gain (posttest – pretest)	5.20 (1.47)	−1.27 (1.49)	$t(19) = 9.99$***	$d = 4.37$
Essays: Correct causes	4.60 (2.98)	4.46 (2.16)	$t(19) = 0.13$	
Reading patterns				
Total time (minutes)	46.96 (20.8)	44.34 (4.2)	$t(19) = 1.05$	
Ratio time				
Reliable/unreliable	3.88 (2.61)	1.39 (.97)	$t(19) = 2.96$**	$d = 1.98$
Unique visits to pages				
Reliable	12.5 (5.72)	8.64 (2.91)	$F(1, 19) = 13.43$**	$\eta_p^2 = 0.41$
Unreliable	8.9 (5.76)	0.09 (2.02)	$F(1, 19) = 1.27$	
Returns to pages				
Reliable	10.20 (5.72)	5.0 (4.52)	$F(1, 19) = 5.88$*	$\eta_p^2 = 0.24$
Unreliable	5.50 (7.56)	7.91 (7.3)	$F(1, 19) = 1.26$	
Difference in ranks				
Reliable – unreliable	2.56 (1.64)	1.97 (1.67)	$t(19) < 1$	

*$p < .05$. **$p < .01$. ***$p < .001$.

sample, this effect was quite small in magnitude and was inconsistent across measures. The main conclusion of Wiley et al.'s study was that the processing and behaviors that readers actually engaged in, regardless of the essay prompt, was a much more robust predictor of learning. In the subsample examined in this study, none of these small effects of the essay manipulation on learning outcomes remained. In fact, of the poorer learners examined here, seven of them had been given the argument prompt, and four received the description prompt. Of the good learners, half came from each essay prompt condition. For all measures considered in this article, a set of parallel analyses was run, including the manipulation factor, and all yielded similar results to the ones reported here. Due to the small sample size of this study and to avoid further reduction of power, the analyses of the think-aloud protocols are reported without this factor.

In summary, the sample of better and poorer learners who are the focus of the present report mimic the major patterns on learning outcome and processing time measures that were obtained in the larger sample reported in experiment 1 of Wiley et al.'s (2009) study.

Processing Measures Derived From the Think-Aloud Protocols

Overall Frequency of Processing Events

The overall frequencies of processing events identified in the think-aloud protocols were submitted to ANOVA. There were significant effects of learner group ($F[1, 19] = 5.56$, $p = .03$, $\eta_p^2 = 0.23$) and reliability ($F[1, 19] = 44.61$, $p < .001$, $\eta_p^2 = 0.70$). However, the significant interaction of the two ($F[1, 19] = 13.98$, $p < .001$, $\eta_p^2 = 0.42$) qualified the main effects: On reliable sites, better learners (M = 164.70, SD = 52.49) generated more processing events than did poorer learners (M = 92.09, SD = 35.69), $F(1, 19) = 19.19$, $p < .001$, $\eta_p^2 = 0.50$. On unreliable sites, the means for better (M = 42.6, SD = 36.5) and poorer learners (M = 57.64, SD = 27.43) were not significantly different, $F(1, 19) < 1$, ns. These data indicate that all learners engaged in more processing on reliable sites but that the magnitude of the difference between reliable and unreliable sites was greater among the better learners.

Relative Frequency of Processing Types

Of more interest is whether the relative distribution of the different kinds of processing varied between the groups and as a function of site reliability. Among the eight first-pass processing categories, three had very low frequencies (prediction, surface connection, and irrelevant association) and were combined to form an "other" category. The resulting six categories are shown in Table 4. To adjust for the overall frequency differences between the learner groups, we computed scores for each individual as the proportion of their total processing events. The overall total includes coded statements for the website with ambiguous reliability (*Volcanolive*). Because better (M = 13.2) and poorer (M = 17.9) learners did not differ significantly on the frequencies coded for this site, the comparisons of proportions are essentially unaffected by using total comments for the seven websites versus the total for the six. We did not further analyze the comments on the *Volcanolive* site.

Descriptively, protocol comments on this site accounted for 5% of the total frequency of protocol comments for better learners and 9% for poorer learners, a nonsignificant difference. For each group, these comments were distributed equally across the processing categories shown in Table 4: For better learners, 0.01 for all but the other category, which was 0; for poorer learners, 0.02 each for all but information/source evaluation, which was 0.01, and other, which was 0. The table provides the mean proportion of events falling into each processing category by learner group and site reliability.

All categories except other met the assumptions for ANOVA according to tests of homogeneity of variance, skewness, and kurtosis. Given the low relative frequency in the other category, we did not analyze data for this category further. Thus, we conducted five ANOVAs for learner group by site reliability, one for each category of processing events shown in Table 4 except for other.

Self-Explanations and Paraphrases

A significant main effect of site reliability was found for self-explanations, with a higher proportion being generated on reliable sites than on unreliable, $F(1, 19) = 43.85$, $p < .001$, $\eta_p^2 = 0.70$. The main effect for learner group was not significant, $F(1, 19) < 1$, ns. However, learner group interacted with reliability, $F(1, 19) = 6.2$, $p = .02$, $\eta_p^2 = 0.25$. Simple effects tests on the interaction indicated that although all learners provided more self-explanations on reliable websites than on unreliable ones, the differential between reliable and unreliable sites was significantly greater for the better learners. Specifically for better learners, self-explanation on reliable sites was significantly greater than on unreliable sites ($F[1, 19] = 44.7$, $p < .001$) with a large effect size ($\eta_p^2 = 0.68$); for poorer learners, it was also significant ($F[1, 19] = 10.0$, $p = .005$), but with a smaller effect size ($\eta_p^2 = 0.32$). A second set of simple effects tests examined the interaction to determine if there were significant learner group differences in the proportions of self-explanations on each type of site. These were for reliable websites but not for unreliable ones. That is, the proportion of self-explanations displayed by better learners on reliable sites was significantly greater than the proportion displayed by poorer learners ($F[1, 19] = 6.0$, $p = .02$, $\eta_p^2 = 0.22$), whereas the groups did not differ on the unreliable sites ($F[1, 19] = 1.67$, $p = .21$).

For paraphrases, a significant main effect of site reliability was also found ($F[1, 19] = 35.63$, $p < .001$, $\eta_p^2 = 0.65$), with a higher proportion of paraphrases

Table 4. Distribution of Comments Coded From Think-Aloud Statements Across Reliable and Unreliable Websites for Better and Poorer Learners

	Better learners		Poorer learners	
Processing category	Reliable M (SD)	Unreliable M (SD)	Reliable M (SD)	Unreliable M (SD)
Self-explanation	0.19 (0.08)	0.02 (0.03)	0.13[a] (0.07)	0.05 (0.03)
Paraphrase	0.16 (0.08)	0.04 (0.04)	0.15 (0.06)	0.07 (0.05)
Monitoring	0.13 (0.04)	0.02 (0.01)	0.07[a] (0.04)	0.03 (0.02)
Evaluation	0.10 (0.05)	0.05 (0.05)	0.08 (0.04)	0.08 (0.04)
Navigation	0.15 (0.04)	0.06 (0.04)	0.10[a] (0.06)	0.08 (0.03)
Other	0.03 (0.02)	0.00 (0.01)	0.04 (0.03)	0.02 (0.01)

Note. Entries in the table are mean proportions for each group of learners. Proportions were calculated for individual subjects by taking the total number of events coded in the particular category and dividing by the total number of events for that individual. Other includes the surface connections, irrelevant associations, and prediction categories, each of which occurred infrequently.
[a]Indicates significant difference ($p < .05$) between mean proportions for better and poorer learners on reliable websites.

being generated on reliable sites than on unreliable ones. However, neither the main effect for learner group ($F[1, 19] < 1$, ns) nor the interaction were significant ($F[1, 19] = 2.14$, $p = .16$). Thus, paraphrasing was distributed similarly across reliable and unreliable sites for both learner groups.

Intertext Connections

To better understand the kinds of connections between ideas that were being made, primarily in the self-explanation comments but also within other kinds of comments, the analyses of intertext connections are considered next. This coding revealed that the frequency of explicit intertext connections was low in general and highly variable across subjects. Descriptive statistics indicated no differences between learner groups (better: M = 5.50, SD = 6.42; poorer: M = 5.20, SD = 3.16). However, the distributions of each learner group departed from normality. The better learners' distribution was bimodal: Three had frequencies of 10 or more intertext connections, but five produced one or none; the remaining two provided three and five connections. Among the poorer learners, the distribution was tighter but again bimodal: Two subjects produced eight intertext connections, two produced one each, three produced three, and the remaining four students produced four, five, six, and 11, respectively.

It was also the case that learner groups did not differ in the likelihood of connecting information across reliable to reliable websites or unreliable to unreliable sites. For reliable site connections, the mean proportion for better learners was 0.40; for poorer learners, it was 0.33. For unreliable site connections, the mean proportion was 0.09 for each group of learners. Thus, for both groups of learners, intertext connections were more likely to connect reliable sites than unreliable ones. However, for both learner groups, intertext connections represented very small proportions of their think-aloud events.

In summary, the analyses of the categories focused on the content-related utterances suggest that even though explicit connections between texts were rare, the better learners more sharply distinguished between reliable sites that were worth investing meaning-making processes in and those sites that had low payoff for building a causal mental model of the Mt. St. Helens eruption. The poorer learners were less discriminating in the allocation of sense-making efforts, working almost as hard to understand information on unreliable as reliable sites.

Monitoring

Main effects of learner group were observed for monitoring, and although significant, the mean differences were small: Monitoring represented a higher proportion of the better learners' processing events (M = 0.075) than it did for poorer learners (M = 0.05), $F(1, 19) = 4.69$, $p = .04$, $\eta_p^2 = 0.20$. There was also a main effect of site reliability for monitoring, $F(1, 19) = 68.19$, $p < .001$, $\eta_p^2 = 0.78$. Furthermore, learner group interacted with reliability, $F(1, 19) = 16.09$, $p < .001$, $\eta_p^2 = 0.46$).

Simple effects tests on the interactions indicated that although all learners provided more monitoring comments on reliable sites than on unreliable ones (better learners: $F[1, 19] = 60.0$, $p < .001$, $\eta_p^2 = 0.79$; poorer learners: $F[1, 19] = 8.0$, $p = .01$, $\eta_p^2 = 0.33$), the differential between reliable and unreliable sites was significantly greater for the better learners. A second set of simple effects tests examined the interaction to determine whether there were significant learner group differences in the proportions of monitoring statements on each type of site. Better learners made significantly more monitoring comments than did poorer learners on reliable sites ($F[1, 19] = 17.0$, $p < .001$, $\eta_p^2 = 0.51$) but not on unreliable sites ($F[1, 19] = 1$, $p = .33$). Interesting, the direction of the means on unreliable sites trends toward the opposite of the direction of the means on the reliable sites. This result is consistent with the summary that emerged from the paraphrase and self-explanation categories.

Information/Source Evaluations

Greater proportions of information and source evaluations occurred on reliable websites than on unreliable ones for all learners, $F(1, 19) = 4.39$, $p = .05$, $\eta_p^2 = 0.19$. There was no main effect for learner group ($F[1, 19] < 1$, ns), nor did the learner group by reliability interaction reach significance ($F[1, 19] = 2.87$, $p = .11$, $\eta_p^2 = 0.13$). Nevertheless, and in part because of the importance of processing information about the source in multiple-text comprehension situations, we did an exploratory content analysis of the information and source evaluation comments. These analyses also allowed us to determine whether the kinds of evaluations made during reading were consistent with the evaluations of the reliability rankings of the websites that were made following the reading, writing, and concept recognition tasks in experiment 1 of Wiley et al.'s (2009) study.

Table 5 shows the subcategories used for classifying the content of the information/source evaluation events, along with examples of each. Consistent with the postreading justifications in Wiley et al.'s (2009) experiment 1, relevance evaluations were the most frequent type of information/source evaluation: They were present in every participant's protocol and accounted for 46% of all evaluations made by poorer learners and 33% of those made by better learners. Evaluations of the quality of the information, including content and reliability/scientific soundness of the information, accounted for 45% of the better learners' evaluations and 30% of the poorer learners'.

Table 5. Information/Source Evaluation Comments for Better and Poorer Learners

Types of evaluations and examples	Better learners Mean = 37.3 (***n*** = 10)		Poorer learners Mean = 29.2 (***n*** = 11)	
	Proportion of total evaluation events = 373	Mean per subject (percentage of subjects providing at least one)	Proportion of total evaluation events = 321	Mean per subject (percentage of subjects providing at least one)
Relevance of the information to addressing why Mt. St. Helens erupted (e.g., "This says nothing about Mt. St. Helens"; "not telling me what I'm looking for"; "not important"; "exactly what I need")	.33	12.4 (100)	.46	13.5 (100)
Reliability/scientific soundness of the information (e.g., "put up by NASA, so it should be reliable"; "no evidence, proof here—pretty far-fetched")	.14	5.4 (80)	.12	3.6 (64)
Source credibility: Author credentials, background (e.g., "It's by *a* professor, so it's probably accurate"; "Just *some guy* wrote it"; "This guy's a crackpot")	.06	2.4 (60)	.01	0.36 (27)
Information or text quality: Substance of the information itself (e.g., "50% isn't a lot"; "That's a lot of information about plants"; "A lot of property was destroyed"), properties of the text (e.g., "good analogy"; "contradicts itself"), physical properties (e.g., "colorful"; "lots of diagrams")	.31	7.0 (100)	.18	5.3 (82)
Affective reaction: Emotive response (e.g., "interesting"; "cool"; "boring"; "weird")	.11	4.2 (60)	.19	5.6 (73)
Search outcome: Success or failure of effort to find something specific (e.g., "not finding the cause here"; "can't find anything about Mt. St. Helens")	.04	1.5 (40)	.02	0.63 (36)

Evaluations of the credibility of the source, specifically of the author of a website, were infrequent in either group of learners, accounting for only 6% of the better learners' evaluation comments and 1% of the poorer learners'. However, 60% of the better learners, as compared with 27% of the poorer learners, evaluated at least one source for the credibility and reliability of the information. This difference is consistent with the more frequent reference to sources in the postreading justifications of the more successful learners in Wiley et al.'s (2009) experiment 1 sample.

Interesting, source credibility tended to emerge during reading in a bottom-up fashion as learners attempted to understand the unreliable websites. Learners did not, except in three cases, use information in the URL of the site to make evaluations of the author or site prior to getting into the site's content. Rather, those mentioning source credibility did so when they evaluated the scientific soundness of the information, calling into question the credibility of the author.

Two examples illustrate this process. Participant 201, a better learner, read on the *Oil Drilling* website that the site's author claimed the discovery of new laws of motion but that the scientific community would not listen to him. Participant 201's comment indicates a connection between this content and the author's lack of credibility among scientists: "Oh, man, she [*sic*] called Isaac Newton a plagiarist. That's probably why they won't listen to her [*sic*]." The second example is also of a better learner reading the *Oil Drilling* site. Participant 204 showed accumulating skepticism of the scientific credibility of the author's argument: The more he read of the information, the more his think-aloud events showed explicit negative evaluations of the content, especially regarding the author's causal conclusions based on correlational data. Participant 204's think-aloud comments also indicated why these conclusions were not scientifically valid. When he left the site, he expressed negative evaluations of the author's credentials and the quality of the evidence, comparing the site and its author unfavorably against one of the reliable sites (*Savage Earth*).

"Who, what is the actual background of this writer? It seems to be most of this evidence is anecdotal. It doesn't seem… he doesn't even use terminology used in the first one, which seemed to be geared towards a, probably a high school or maybe even grade school type reader: giving terminology and explaining what the mechanisms of volcanoes…This is…This is a politically slanted webpage."

These source evaluations emerged as readers grappled with the content of the websites and in connection with assessments of the quality of the evidence on the sites. Thus, although there were no significant differences between the learner groups in number of explicit information/source evaluations that were stated in the protocols, the content analysis is consistent with predictions that better learners would be more sensitive to issues of scientific soundness (reliability, validity) of sites and author credibility.

Navigation

The comments related to navigation showed a significant main effect of reliability ($F[1,19] = 17.10$, $p < .001$, $\eta_p^2 = 0.47$) that was qualified by a significant interaction with learner group ($F[1, 19] = 9.07$, $p = .007$, $\eta_p^2 = 0.32$). The main effect for learner group was not significant, $F(1, 19) < 1$, ns. Simple effects tests on the interaction indicated that only the better learners showed a significant difference between reliable and unreliable sites, $F(1, 19) = 22.0$, $p < .001$, $\eta_p^2 = 0.56$. Poorer learners did not differentiate, $F(1, 19) < 1$, *ns*. Furthermore, better learners provided significantly more navigation statements on reliable sites than poorer learners did ($F[1, 19] = 7.5$, $p = .01$, $\eta_p^2 = 0.31$), but there was no significant difference between the groups on unreliable sites ($F[1, 19] = 1.5$, $p = .24$). Thus, better learners were more likely to verbalize about their moves from page to page and from site to site than were poorer learners only when they were on reliable sites.

Content analyses of the navigation events provide more direct evidence of differences between the learner groups because these comments conveyed information regarding learners' search and selection goals, what they were looking for as they considered which sites to enter, what to read versus skim over on a page, why they were reading sections versus skimming them, and reasons for continuing to read a site or go to another. We focused our content analyses on the reasons subjects reported for navigating to and away from a page they were currently reading. Both learner groups reported similar reasons for going to reliable and unreliable sites: The overwhelmingly dominant reason was the presence in the Google search page title or tagline of a keyword related to the task (e.g., "causes of volcanoes," "Mt. St. Helens," "volcanoes").

Where the learner groups differed was on reasons for leaving reliable and unreliable webpages. Three subcategories (new goal, got to end of page, irrelevant information) captured the majority of the reasons given for leaving pages. Their frequencies are shown in Table 6. The new-goal subcategory violated the assumption of homogeneity of variance, and several other subcategories violated the assumption of normality. Accordingly, we conducted a series of nonparametric Wilcoxon signed-rank tests comparing navigation reasons on reliable and unreliable webpages for each of the three subcategories for better learners and again for poorer learners. These tests demonstrated that better learners were more likely to leave reliable webpages compared with unreliable ones because they had new goals ($W = 1$, $Z = -2.31$, $p = .02$, $r = -.73$) and because they reached the end of the page ($W = 1$, $Z = -2.67$, $p = .007$, $r = -.84$). Conversely, better learners were more likely to leave unreliable pages compared with reliable ones because of the irrelevancy of information ($W = 1$, $Z = -2.31$, $p = .02$, $r = -.73$).

The pattern of effects for the poorer learners showed far less differentiation in reasons for leaving pages on the reliable websites as compared with unreliable ones. Indeed, the only reliable difference was for irrelevance of information: Poorer learners offered this as a reason for leaving pages on unreliable sites more often than as a reason for leaving pages on reliable sites

Table 6. Navigation Comments: Mean Frequencies of Different Reasons for Leaving Pages on Reliable and Unreliable Websites for Better and Poorer Learners

	Better learners		Poorer learners	
Reason for leaving pages	Reliable M (SD)	Unreliable M (SD)	Reliable M (SD)	Unreliable M (SD)
New goal	6.70 (3.97)	2.20[a] (2.44)	2.09 (2.07)	2.73 (2.45)
Got to end of page	4.40 (2.27)	0.80[a] (0.92)	3.55 (2.95)	1.64 (1.50)
Irrelevant information	1.20 (0.79)	5.40[a] (4.12)	2.64 (2.84)	4.91[b] (3.11)

[a]For better learners, reliable and unreliable differed significantly ($p < .05$).
[b]For poorer learners, reliable and unreliable differed significantly ($p < .05$).

($W = 1$, $Z = -2.61$, $p = .009$, $r = -.79$). The remaining subcategories did not show significant differences ($W = 1$, Zs < -1.61, ns). Thus, the poorer learners were equally likely to develop new goals upon leaving unreliable and reliable sites and were just as likely to read to the end of pages on unreliable and reliable sites. The differences between the better and poorer learners' reasons-for-leaving profiles are consistent with the patterns of differences in allocation of sense-making (i.e., self-explanations) and monitoring processes.

Essentially, better learners left what they were reading on reliable webpages because they had goals to support or elaborate their current understandings with additional information. Sometimes they reported that something on the current page suggested other information that they wanted to go and look for or reminded them of information that they had read previously and wanted to return to, presumably in service of reviewing the information. Other times, they reported that they had gotten everything they could from the current page and needed additional information. Thus, especially when leaving pages on reliable sites, better learners were significantly more likely than poorer learners to have completely read the pages and specified goals for the next step in their reading process. Although poorer learners indicated awareness of the irrelevance of information on unreliable webpages (a pattern that better learners also displayed), poorer learners often stated this as simply the absence of keywords having to do with volcanoes or Mt. St. Helens.

Summary From Processing Analyses

The analyses of the relative frequencies and content of the different processing event categories indicate that better and poorer learners generally did not differ in which processes they used, but rather in when they chose to use them: Better learners showed a larger differential preference to employ self-explanation and monitoring on reliable sites. Better learners' reasons for leaving pages reflected greater planfulness and goal-directedness than those of the poorer learners, especially on reliable sites. Finally, the information/source evaluation comments suggest a greater tendency among the better learners to take note of information quality and credibility than the poorer learners did, while relevance seemed the primary driver of the information/source evaluations of the poorer learners.

Contrastive Case Studies

What is not reflected in these separate analyses of the different processing categories is the dynamic nature of the reading process and the ways in which the various kinds of processing activities were assembled and contributed to readers' decisions about how to explore the websites to gather information that would help them address the inquiry question. As we analyzed the protocols of the better and poorer learners for purposes of categorizing each think-aloud event, it became clear that there were important differences among the better and poorer learners in how they were building toward particular decisions reflected in their navigation behavior.

Although some of the differences are reflected in the content analyses of the individual categories of processing activities, much is not evident. The low sample size precluded taking any sort of computational approach to understanding the patterns and sequences of processing activities that we were noticing. Instead, we adopted a case study approach similar to approaches taken by Hartman (1995) and Coiro and Dobler (2007) in their studies of multiple-text reading and Internet reading, respectively. We focused the cases on processing leading up to navigation decisions and actions, including website selection for reading, navigation within sites, and navigation to different sites.

We selected four representative cases, two of better learners and two of poorer learners, based on think-aloud comments in the navigation category. All four had individual percentages of their total think-aloud events that reflected the mean performance of their respective learner groups on navigation events for reliable and unreliable sites. For each, we provide a trace of their reading to illustrate their cognitive processes surrounding navigation within a site and between sites. In general, the cases provide additional evidence of how better and poorer learners differed in their approach to the multiple-text inquiry task.

The two representative better learners were similar in that navigation decisions were related to determining whether what they were reading was providing the kind of information that they were looking for. They differed in terms of the specificity of their goals: Participant 225 had a more specific goal (finding information on causes of volcanic eruptions) than did participant 230, whose goal was more general (information on Mt. St. Helens). The two poor learner cases (participants 236 and 217) provide a contrast with the better learners in that the two poorer learners did not appear to base their allocation of research time on the information that was most relevant to the inquiry task.

Better Learner 225

Participant 225 established a consistent pattern of processing activities. He began by indicating what he was looking for (navigation—goal) and monitoring his state of knowledge about Mt. St. Helens. As he got into the content of the websites, he engaged in efforts to explain and elaborate on what he was reading. Segment 1 in Table 7 presents his sequence for the first site that he read. In line 1, he established that his goal was to

Table 7. Think-Aloud Protocol Segments for Better Learner 225

Think-Aloud Verbalizations[a]	Processing Category
Segment 1	
1. Okay…alright well…first I want to find out a little more maybe about Mt. St. Helens.	Navigation (goal)
2. I really don't know much about it…I've heard… I've heard about it.	Monitoring (prior knowledge)
3. So let me just look at my options…and I'm seeing	Navigation
4. *Volcano* ***[reads title for*** **Volcano*, the NASA site]***	
5. So I'm gonna click on *Volcano*. ***[clicks URL and enters*** **Volcano*, the NASA site]***	Navigation
6. The first thing I see is the map and just to kind of get the location or just to find out where Mt. St. Helen's was located,	Navigation (goal)
7. 'cuz I really don't know so I don't know.	Monitoring
8. I'm just gonna read and see what it says.	Navigation
9. ***[reads first several sentences****]* *Imagine taking a world map, closing your eyes and putting your finger down on the map anywhere at random. If you were instantly transported to that spot on Earth and were to look around, do you think you would be able to see a volcano? Or even more exciting, would you see an erupting volcano? Probably not, because most volcanoes, especially active ones, occur in only a few well-defined narrow bands across the face of Earth.*	
10. Makes sense.	Monitoring
11. ***[reads next sentence****]* *Why do most volcanoes* ***[skips word*** **occur]** *in designated in designated narrow bands?* ***[repeats phrase "in designated"]***	
12. I don't know what it's talking about. ***[giggles]*** I don't know what a narrow band is.	Monitoring
13. Umm…let's see what it says.	Navigation
14. ***[reads next sentence]*** *Why not everywhere, such as in your backyard?*	
15. Well, 'cuz volcanoes don't occur just anywhere.	Self-explanation
16. ***[reads next sentence]*** *Why are some explosive and some not?*	
17. Well, I think it has to do with something in the ground that ruptures the, umm, the particles probably in a volcano.	Self-explanation
18. ***[reads next sentence]*** *For that matter, why do volcanoes occur at all?*	
19. Probably different frequencies in the ground or something	Self-explanation
20. I don't know.	Monitoring
21. ***[reads next sentence]*** *Reasonable understanding of the answers to these questions has only been attained during the last hundred years or so.*	
22. OK**…**all right…I'm just gonna skim a little to see maybe where the volcano is located or where different volcanoes are located…and it says	Navigation (goal)
23. ***[reads next sentence]*** *An important clue to understanding volcanoes is knowing the location of the volcanic bands.*	
24. That's a…I don't know what a…I'm guessing a band.	Monitoring (meaning)
25. Its probably got to be…maybe stretches over a certain area like a rubber band. It could stretch far, so I'm guessing volcanic bands stretches over, umm, some particular area throughout the world.	Self-explanation
26. ***[reads next sentence]*** *Many of the world's active volcanoes are located around the edges of the Pacific Ocean: the west coast of the Americas, the east coast of Siberia, Japan, the Philippines, Indonesia, and in island chains from New Guinea to New Zealand—the so-called "Ring of Fire."*	
27. OK…Yeah, I can, I can see now.	Monitoring

(*continued*)

Table 7. Think-Aloud Protocol Segments for Better Learner 225 (*continued*)

Think-Aloud Verbalizations[a]	Processing Category
28. Where the volcanic band, or as it calls the ring of fire, I'm guessing, uhh, volcanoes are located all around the Pacific Ocean and the bordering continents. That's my guess.	Paraphrase (perhaps to state his understanding)
29. Umm…I'll just read on to find out what else this has to say.	Navigation (continue reading on page)
Segment 2	
1. Just find out more ***[clicks to return to the Google search menu]***	Navigation
2. Hmmm…I don't understand it through the page yet…you know, I kinda understand that.	Monitoring
3. I learned about the plate tectonics and learned that they…a volcano can be more explosive when the…when the plates are more, uhh, compressed or more together, and then as the article said with Ireland and Hawaii, they are more spread apart.	Self-explanation
4. So I want to find something more…more about, uhh, Mt. St. Helens. So maybe read about it as, umm…	Navigation (goal)
5. ***[clicks on the* A Blast From the Past *website and goes to page 1, reading from the top of the page]*** *A Blast From the Past*	(Unreliable)

Note. Italicized text is information being read. Learner comments are in regular font. Coder annotations are boldface in brackets. Coder comments on the processing category appear in parentheses.

[a]The texts being read are from the *Volcano* (www.nasa.gov/volcano.html) and A Blast From the Past (www.stariq.com/blast.html) websites.

find out more about Mt. St. Helens because he did not know much about it. He selected the first site listed, the reliable site *Volcanoes*. He saw a map and stated a new goal of finding out the location of Mt. St. Helens because he did not know where it is.

He then read several sentences, indicating that it "makes sense" (line 10). He read the next sentence, indicated that he did not understand it, and decided to see what else the text says (line 13), perhaps hoping that the text would clarify. He read and self-explained the next several sentences, continuing to question his understanding (line 20). He continued reading, and in line 22, he decided to keep going, continuing to monitor and self-explain. At line 27, he indicated that he now understood, and he paraphrased the previous sentence, perhaps to make visible his understanding. Having resolved his confusion, he went on to read to the end of the page, continuing to use a pattern of making monitoring statements and providing explanations and sometimes paraphrases of what he was trying to understand. When he got to the bottom of the page, he stated a new goal of finding out why volcanoes erupt. He went back to the Google page, indicating that he did not find out why from the page that he was reading, and provided a self-explanation of what he learned on that page and his decision to go to another site. This sequence is shown in segment 2 of Table 7.

The second website that he went to was unreliable (*A Blast From the Past*). However, the first page of the site provides descriptive information about the magnitude of the eruption and the effects of the blast. Throughout this page, participant 225 used the same pattern of monitoring his understanding, self-explaining, occasionally paraphrasing what he was reading, and continuing to read. At the end of this page, he stated that he was not getting information about the causes and decided to leave and seek the information on another site: "I want to find more maybe about the causes of the eruption. Let's see what other options there are."

The third website listed on the Google search results page has a tagline that explicitly discusses causes of volcanic eruptions ("The third process that causes volcanic eruptions is an injection of new magma into a chamber that is already filled"), and participant 225 decided to "check this out." He used the same pattern of monitoring, self-explaining, and continuing to read until, at the end of page 1, he understood the information on this site about the buildup of pressure in the magma chamber, as indicated by his self-explanations and monitoring statements. At that point, he decided to go to another site to see "what other information I can gather."

Finding out more about the causes of volcanic eruptions continued to be participant 225's main goal as he selected other websites and went through them. On each site, he maintained the pattern of monitoring his understanding by providing self-explanations and paraphrases along with navigation decisions. By the end of the research phase of the study, participant 225 had read through at least the first page of all but one of the sites and had read several pages on both reliable and unreliable sites. He rarely evaluated the information, and when he did, it was in terms of what it had to do with causes of volcanic eruptions.

Better Learner 230

Participant 230 used a combination of scanning and deeper reading strategies depending on whether he thought the information was relevant or seemed relevant to the Mt. St. Helens event. When it did, he engaged in self-explanations and comprehension-monitoring processes. Otherwise, he scanned the pages, looking at most of the pages on the website. Within the first minutes of the research phase, participant 225 demonstrated sensitivity to quality of information. He began by describing his navigation activity: "I'm just scanning the websites looking at the addresses to see if they're…umm…like credited sites or if they're just like weird addresses."

He then entered the first site listed on the Google search results page, the reliable site *Savage Earth*. He scanned over the first paragraph, which he described as "more like a story," perhaps based on the fact that it is written to draw the reader in. He started to read sentences aloud with the second paragraph and proceeded to indicate whether he knew the information in the text already and to generate self-explanation comments. At the end of the first page, he indicated that he would go on to the next page. He read the first paragraph but skipped over the second one "'cause it's just quotes and stuff." He resumed the pattern from page 1 for the remainder of the page: read a sentence aloud, generate self-explanation and/or monitoring comments, and add an occasional evaluation that the content was relevant to Mt. St. Helens. He decided to go to the third page of the site, "just scanning the page for anything about St. Helens or any other dates." He indicated that he found nothing and was leaving the site to find something about Mt. St. Helens.

He scanned the titles and tagline on the Google page and went to the last one listed, *A Blast From the Past* (unreliable) "'cause it's about St. Helens, although… (clicking into the website) the website doesn't seem real. I'm just gonna read through." The content on the first page of this site is factual, but he scanned it "'cause it's just the background information." He continued reading sentences aloud to the bottom of the first page and went on to the second even though he stated that "(it) seem(s) to be just talking about…the site of it…not the cause of it." After reading the first couple sentences on page 2, which discuss astrologers making charts for people, he scanned several additional sentences and evaluated the credibility of the site: "I'm scanning, and I see things like Scorpio and Taurus and Virgo, and it seems kind of like this is gonna be B.S.…OK, yeah, I really don't wanna read this, but I'll go to page 3."

He moved the pointer through the page, concluded that it was about the background and was not that important, and left the unreliable site without reading the rest of it. This reading behavior of scanning and leaving without actually reading the sentences aloud was one strategy that better learners used on the unreliable sites. (See convergent eye-tracking results for a different subsample of the Wiley et al., 2009, study reported in Wiley et al., 2011).

The switch between scanning and reading sentences aloud was a consistent pattern that participant 230 used throughout the first third of his research phase. From the *A Blast From the Past* website, he went to the unreliable *Danger Days* site next and scanned through its pages "looking for anything that talks about St. Helens." Evaluating the site as not having anything relevant to Mt. St. Helens, he went back to the Google page without actually reading aloud any of the *Danger Days* text. He found some relevant information on the *Volcanolive* site, read this aloud, and provided self-explanation and monitoring comments for these sections. When he got to page 2, he indicated that the topic had shifted (to Hawaiian volcanic eruptions); he changed to a scanning strategy and reached the conclusion that there was nothing relevant.

He went to the *Volcano* site next and spent the remainder of his research time there. Interesting, he also provided one of the few source evaluations that were generated during reading: "There's a NASA logo in the corner, so it seems this might be [a] more credited site." While on this site, participant 230 read aloud, and for almost every sentence or couple of sentences, he self-explained with an occasional paraphrase and monitored his understanding. Interesting, he also followed a number of the links to diagrams and maps that were part of this site and connected the information in those to the verbal information that he was reading. As he read more on this site, his self-explanations increasingly attempted to relate the different aspects of a volcanic system. He also provided further evidence of the strategic nature of his reading processes when he came to a section of text that he was having some trouble understanding: "I'm just trying to read it slower to kind of get an idea of, you know, a more concrete idea of what's going on."

The overall patterns for each of the better learners illustrate purposeful allocation of processing to information that these individuals evaluated as relevant to meeting the task goal of producing an essay about the cause of the Mt. St. Helens eruption. These participants made decisions about whether to leave a site or keep reading based on the success of efforts to make meaning from text that they evaluated as relevant to the task. This pattern in the better learner group as a whole is consistent with the differential time allocation to reliable websites, as compared with unreliable ones, and the differential occurrence of self-explaining and monitoring on reliable websites, as compared with unreliable ones.

Poorer Learner 236

Participant 236 demonstrated a consistent pattern of paraphrasing information, an occasional self-explanation, relevance evaluations, and navigation comments that tended to describe what she was doing but not why. She began by reading the titles and taglines on the Google search results page, "just to see what they're about." After reading through the list, she said she was going to the first one, the reliable site *Savage Earth*. After reading the first three sentences, she indicated that she was skimming and provided affective and relevance evaluations: "Volcanoes are very interesting, but it doesn't say anything about what caused Mt. St. Helens to erupt." She started reading again in the third paragraph on page 1, paraphrased it, and left the site to go to the second site listed "to see if it says anything about Mt. St. Helen." Her navigation decision, although goal based, was nonspecific to what she wanted to find out about Mt. St. Helens.

She went to the *Oil Drilling* site (unreliable) and read it to the end of the first page, making a couple paraphrases ("This just talked about Iran and Turkey"), relevance evaluations ("This has nothing to do with what I am looking for"; "That's nothing"), and information-type evaluations ("sounds like it's a report, its like statistics") before she left the site. The next site that she went to was the reliable *Scientific American* site. The pattern of reading several sentences with a few scattered paraphrases of the content information continued on this site. Interesting, although participant 236 evaluated the information on this page, she appeared to dismiss it as irrelevant because it was about volcanoes in general and not specifically about Mt. St. Helens, stating, "That's about all volcanoes in general," right before leaving the page. This is ironic in that the volcanic processes that she had read about are part of the causal model for the Mt. St. Helens eruption.

Participant 236 next went to the unreliable *Danger Days* site, decided that it was relevant, and proceeded to again read and paraphrase groups of sentences, with an occasional simple self-explanation (e.g., "So it [tidal force and sunspots] will probably be [lead to] more earthquakes"). She continued to page 2, skimmed and noted that it did not mention Mt. St. Helens, and went to the Google page. She then entered the *Volcanic Eruptions and Tides* site, read one page, returned to *Danger Days* briefly, and then went to the reliable *Volcano* site. She read all the pages on the *Volcano* site but, as with the other sites, read several sentences at a time, offering perfunctory paraphrases of the information and indicating that "makes sense." The last website that she read was the unreliable *A Blast From the Past* site, where she read all the pages. Having "been to" all the sites, she decided that she was done researching after 25 minutes, stating, "I think I got information, not necessarily on Mt. St. Helens itself but on, like, all volcanoes, so that's fine." She then proceeded to the writing task.

Participant 236 provides an interesting example of what might be called illusory understanding. Her pretest to posttest volcano concepts performance indicated no change in what she understood about the causes of volcanic eruptions. Her decisions to leave websites seemed to be procedurally based rather than conceptually based in that she seemed content with just restating what she had read in similar words. Having read through each page, she was ready to move on. Although representative of poorer learners with respect to her navigation events, completing the research phase in 25 minutes was not representative of the other poorer learners, all of whom used the entire research time.

Poorer Learner 217

Participant 217 showed little discrimination among task-relevant and task-irrelevant websites and information within them. He engaged in paraphrasing material that he read aloud or self-explanations that tended to expand only on the sentence that he had just read. He did not use prior knowledge or other information from other sentences that he had already read in his explanations. Participant 217 provided almost no evaluative comments and infrequently monitored his understanding. Indeed, the evaluations that he made led him to dismiss important information. For example, at the start of the research phase, he indicated that he was "passing up the first one [*Volcano*] because it's kinda basic."

He went to the next one on the Google search results page, the unreliable *A Blast From the Past*. After reading the first two sentences on this site, he provided a self-explanation that tried to unpack what the text meant with the phrase "the names 'Washington' and 'Helen' were literally written in the stars when the volcano roared to life." He continued reading aloud and paraphrasing to the bottom of the page and went on to the next "to see what page 2 is about." Although he provided self-explanations for three of the sections of text on page 2, they were vague about the relation to the eruption: "It was kind of Mars and Uranus, you know, that had something to do with the eruption." He went on to page 3 with no indication of why and proceeded to read aloud and paraphrase the sentences on this page. He went to page 4 and provided a self-explanation for information about the recovery of the area devastated by Mt. St. Helens. Interesting, he read information about the author that appeared on this last page of the site but made no comment about her qualifications with respect to providing an explanation for the Mt. St. Helens eruption.

He left the site and returned to the Google page, where he read the titles and taglines for three sites:

> "*Volcano* - An important clue to understanding volcanoes is knowing the location of the volcanic bands..."; "*Scientific American: Ask the Experts* - ...The third process that causes volcanic eruptions is an injection of new magma into a chamber that is already filled...; *Volcanic Eruptions and Tides* - ...reputable scientists have suggested that tides might influence whether a volcano will erupt or not...."

He then went to the *Volcanic Eruptions and Tides* website, saying that he wanted to "see if it had something to do…." After reading the first several sentences, participant 217 decided that the tides probably did have something to do with the eruption, and he kept reading. When he read, "it should not come as a complete shock that reputable scientists have suggested that these squeezings [described in previous sentence] might influence whether a volcano will erupt or not," he indicated that he was having trouble understanding this idea. The last sentence that he read on this page indicates that "volcanoes are more likely to erupt at the fortnightly (or 14 day) 'high' tide."

He went on to read the second page that stated the following:

> Nearly twice as many eruptions have occurred nearer fortnightly tidal maximum than tidal minimum. Hawaiian Volcano Observatory scientists have noted that the Pu'u 'O'o fountaining episodes each occurred remarkably close to fortnightly tidal maximums and that the first set of eruption pauses in 1990 (periods during which the eruption turned off for up to a few days) occurred remarkably close to fortnightly tidal minimums.

From the co-occurrence data of tides and eruption episodes, participant 217 concluded that "that's probably what caused some of it [eruptions] with the tides." Thus, he appears to have come away from this site with the moon and tides as causes of volcanic eruptions, however imprecise his understanding. Although he moved to the next page of this site, he read only the first sentence and then went back to the Google page with no indication of why.

Back on the Google page, he read the titles of *Danger Days* and the site that attributes volcanic eruptions to oil drilling, *The Cause of Most Earthquakes, Volcanoes and Bad Weather*, and went to that site. He read through two pages, providing paraphrases and self-explanations that related two segments of the text, neither of which discussed volcanic eruptions. He then read the explanation provided on this site for volcanic eruptions: Pumping too much oil creates conditions that affect barometric pressure, which causes the earth to wobble, "and if it continues to get worse it could cause the ocean tides to get very dangerous or even worse create a worldwide earthquake or volcanic eruption!" He generated a self-explanation connecting earlier statements about earthquakes in Iraq and Turkey to pumping oil, apparently accepting the causal explanation offered by this site. He continued to the end of this site, reading and providing paraphrases.

Then, participant 217 went to the reliable *Volcano* site and used the causal model that he had developed from the sites read thus far in his efforts to understand the information on the *Volcano* site. He put together information about volcanoes that occur in the ocean with the tides information even though Mt. St. Helens is not an oceanic volcano. Thus, he emphasized in his reading and self-explanations ideas and concepts that were not relevant to the task or that were not corroborated by the reliable sites. He did eventually get to information about subduction and Mt. St. Helens, but he left the research with only a vague understanding of causes of volcanic eruptions: "plates colliding…dipping beneath each other…what did they [the text] say—heat up or something. The plates dives into the mantle—the layer of hot…uh, I guess that's just what does it—the plate gets hot."

In brief, then, participant 217 spent about half of his research time reading unreliable sites that contributed to his developing a causal model, albeit vague, of volcanic eruption related to tides and the impact of oil drilling and pumping on the stability of the earth. He spent the other half of his time on reliable sites, but his processing and learning from these may have been negatively impacted by the understanding acquired from the earlier reading. Some support for this comes from his postreading essay, in which he included pieces of both the correct and incorrect causal models.

There are clear contrasts between the two poorer learners in that the first did not invest very much effort in the task, and the second, although spending the whole research time reading, paraphrasing, or self-explaining, with some monitoring, did not appear to relate these processes to each other or toward navigation decisions. Nor did he attempt to apply criteria of relevance or source quality. He ended up spending as much effort on faulty models as accurate ones, despite the fact that he was not different from other subjects in educational background in science.

Summary From Contrastive Cases

The picture that emerges from these contrastive case analyses is that the better learners were more strategic than the poorer learners in both how and what they read. Better learners used more monitoring and evaluation processes to determine not only what they understood from the information provided but also whether it was scientifically credible or task relevant. Although decisions to visit sites were made on a similar basis—using keywords to infer the presence of information likely to be task relevant—the patterns of processing events indicated more evaluation on the part of the better learners regarding how the information that they were looking at did or did not further their understanding of volcanoes in general and the eruption of Mt. St. Helens specifically. Decisions to continue on or seek information on other sites followed from these evaluations.

General Discussion

The primary focus of this report is on using think-aloud protocols as a vehicle for understanding the processing activities of better learners, as compared with poorer learners, when conducting a Web-based science inquiry task. In this sample, as in the larger sample reported in experiment 1 of Wiley et al.'s (2009) study, the students who learned more spent more time on reliable websites relative to unreliable ones than did those students who learned less. Although all learners spent more time on reliable sites than unreliable sites, the better learners showed a substantially larger preference for reliable sites. The distribution of processing events critical for sense making, namely self-explanation, showed this same differential bias between better and poorer learners (even after accounting for overall differences in the amount of think-aloud comments). The same patterns were seen for monitoring. That is, both groups of learners engaged in more of these kinds of processing activities on reliable sites, but the difference between reliable and unreliable sites was substantially greater among the better learners. The differential was not present for other kinds of processing activities that contribute to establishing the literal meaning of the text (e.g., paraphrasing).

Some of the most striking differences between better learners and poorer learners were in the content of their information/source evaluation and navigation comments. Analyses of the content of the evaluations indicated that better learners were more likely to evaluate source information for credibility and reliability. That is, remarks about how they were making decisions about which sites to read, whether to continue reading, or when and for what reasons they would leave a site revealed that the better learners were more strategic in terms of what they selected to read. Their more strategic approach was largely based on awareness of their task, their current understanding, and assessment of what they still needed to accomplish the task. Consequently, the navigation decisions of the better learners were more explicit and goal-directed: They incompletely read pages that they judged would not further their understanding and finished pages that furthered their understanding, leaving them with new goals in mind.

The question remains, How did the better learners know which sites to spend more time on? Given the overlap in pretest scores between better and poorer learners, it does not seem that the observed processing differences are attributable to preexisting differences in knowledge of volcanoes or science background, at least as measured by the volcano concepts pretest and number of science courses. The evidence available from this study does not support an explanation based on better learners having a priori knowledge of which sites might be considered more reliable and useful, in the way that disciplinary experts might.

While experts may be able to employ top-down evaluation processes that may allow them to select more reliable sources for consideration early in a research process, the evaluation processes observed in this sample seemed to emerge as a result of meaning-making processes. There were few statements such as "This is a NASA site, so it should be good," or "This is a dot-com site, so I might not be able to trust what it is telling me." Source evaluation based on features of the sites present on the Google search results page was simply not a top-down driver of what sites to visit, consistent with observations from other studies (Braasch et al., 2009; Brem, Russell, & Weems, 2001; Kiili, Laurinen, & Marttunen, 2008; Kuiper et al., 2005; Walraven et al., 2009). Rather, the present data suggest that it was actually the explanation-driven processing while reading the websites, in conjunction with an awareness and evaluation of whether understanding was increasing and what additional information was needed, that resulted in better learners spending a greater proportion of their time on the more reliable sites than did the poorer learners.

Thus, these processing activities serve as a means for interpreting Wiley et al.'s (2009) findings that comprehension and evaluation behaviors are closely linked when learning from Web-based inquiry tasks. More work is needed to better understand how manipulating features of document sets or inquiry tasks, and how individual differences in constructs such as epistemic beliefs, might alter the kinds of processes that readers use, and how they use them, during learning from multiple sources (cf. Ferguson, Bråten, & Strømsø, 2012; Mason, Ariasi, & Boldrin, 2011; Mason, Boldrin, & Ariasi, 2010; Thomm & Bromme, 2012).

One possible concern about this sample was the failure to find writing prompt effects. This writing task manipulation consisted of a single word change, and interpretation was left up to the students. When the upcoming writing task was referred to by students in their think-aloud comments, it was in terms of finding causes for volcanic eruptions, and the students did not mention the format of the essay that they were preparing for. All of this suggests that the writing prompt manipulation was quite subtle, and its effects may be imperceptible especially when one is examining the reading and thinking that comes before writing the essay. These effects were weak even in the larger sample. As in the larger study, the detailed analysis of this sample suggests that the processing that students engaged in during reading was more important than the essay prompt manipulation for determining performance.

Another concern about findings from this sample could be due to the use of examples in think-aloud instruction. By providing a range of examples, we endeavored not to bias readers toward deeper processing but rather just to increase their awareness of several kinds of comments that they could make. Yet, we acknowledge that the specifics of a think-aloud instruction can affect the likelihood that students engage in evaluative processing in multiple-text situations (Gerjets, Kammerer, & Werner, 2011). There is some evidence that think-aloud protocols that focus on self-explanation may improve comprehension and learning (Chi, 1997).

However, intervention studies also show that simply telling people to use explanation and integration processes often does not result in people actually using them (McNamara, 2004). More extensive training, including modeling and practice with feedback, is typically necessary to achieve even small effects (McNamara, 2004). Thus, we think it unlikely that including a range of example think-aloud comments alongside a practice text as part of instruction was responsible for the patterns of processing seen here. All participants received the same instruction, yet differences were seen between better and poorer learners.

A somewhat unexpected observation was the low rate of comments that reflected intertext connections during reading even among the better learners. This was unexpected based on two of the learning outcomes. First, items on the volcano concepts recognition test were created to reflect connection making. Because performance gains on this test were the basis for defining the better and poorer learner groups, it was surprising that the better learners did not make more comments that reflected intertext connections. Second, at least half of the better learners, but only one of the poorer learners, produced essays that showed evidence of an integrated causal model for the Mt. St. Helens eruption. On reflection, however, the seeming discrepancy between the content of the think-alouds and performance on the learning outcome measures may be related to when such integration occurred. The present data provide scant evidence of it occurring during reading, suggesting that it occurred during essay writing and/or responding to the volcano concepts assessment. Additional research is needed to better understand how and when readers make connections across information occurring in different sources as well as within the same source.

Several important conclusions are possible based on the presented findings. First, simple counts of instances of one type of processing versus another did not discriminate among better and poorer learners. Instead, it was when these processes were employed, in what contexts, and in what order or combination that mattered. The cases provide examples of planful navigation decisions that were associated with effective learning. These learners were both aware of and attempting to make relationships among bits of information, whether that information was distributed across documents or was prior knowledge that the students brought into the learning situation. The present findings show that in multiple-text situations, self-explanations not only are critical for constructing relatively complete and coherent conceptual models but also play an integral role in evaluating whether sites that are selected on the basis of surface overlap of keywords and phrases with an inquiry task actually contain information relevant to completing the task. This conclusion extends findings from single-text studies about the facilitative relationship between better comprehension and self-explanation (Ainsworth & Burcham, 2007; Chi et al., 1994; Coté et al., 1998; Ozuru, Briner, Best, & McNamara, 2010).

Furthermore, the present work demonstrates the important role of comprehension monitoring in multiple-text situations and how engaging jointly in both self-explanation and monitoring can contribute to comprehension. In the present data, monitoring in the absence of self-explanation was often very superficial (e.g., "I get that"; "I don't get that") without the articulation of next steps. In the better learners, self-explanations served as a catalyst for monitoring, consistent with suggestions that self-explanations may improve reading by bringing comprehension processes to the explicit attention of the learner (Griffin, Wiley, & Thiede, 2008; McNamara & Magliano, 2009). When self-explanation and comprehension monitoring were used in concert, and when these processes were used to inform evaluation and navigation decisions, learners were more strategic in their subsequent comprehension and learning efforts.

Thus, there is an intimate and dynamic relationship between meaning making, comprehension monitoring, and source evaluation in the context of self-regulated learning activities such as Web-based inquiry tasks (Griffin et al., in press; Winne & Hadwin, 1998). The processing picture that emerged from the contrastive analysis of the think-alouds of the better and poorer learners in this study is consistent with a conception of reading as a dynamic process (Goldman & Saul, 1990; McNamara & Magliano, 2009).

A dynamic model of reading to learn, especially from multiple sources of information, implies that reading comprehension instruction needs to emphasize much more than introducing different strategies to readers and having them practice each, in relative isolation from one another. Rather, instructional models and design principles need to attend to the complex interrelationships between meaning-making, evaluation, and monitoring processes as students attempt to develop understanding by reading multiple sources.

Notes

The work reported in this article was supported, in part, by Grant REC 0126265 ("Understanding in Science") from the National Science Foundation to Jennifer Wiley, Susan R. Goldman, and Arthur C. Graesser. The authors thank Ivan K. Ash, Rona Gepstein, Joshua Hemmerich, Melinda Jensen, and Christopher A. Sanchez for their contributions to this project. The preparation of this article was also supported by the Institute for Education Sciences, U.S. Department of Education through Grant R305F100007 to the University of Illinois at Chicago, Susan R. Goldman, PI, and Grant R305B07460 to the University of Illinois at Chicago, Jennifer Wiley, PI. The opinions expressed are those of the authors and do not represent views of the funding agencies.

References

Ainsworth, S., & Burcham, S. (2007). The impact of text coherence on learning by self-explanation. *Learning and Instruction, 17*(3), 286–303. doi:10.1016/j.learninstruc.2007.02.004

Azevedo, R., & Witherspoon, A.M. (2009). Self-regulated learning with hypermedia. In D.J. Hacker, J. Dunlosky, & A.C. Graesser (Eds.), *Handbook of metacognition in education* (pp. 319–339). New York: Routledge.

Bazerman, C. (1985). Physicists reading physics: Schema-laden purposes and purpose-laden schema. *Written Communication, 2*(1), 3–23. doi:10.1177/0741088385002001001

Berkenkotter, C., & Huckin, T.N. (1995). *Genre knowledge in disciplinary communication: Cognition/culture/power.* Hillsdale, NJ: Erlbaum.

Braasch, J.L.G., Lawless, K.A., Goldman, S.R., Manning, F.H., Gomez, K.W., & MacLeod, S.M. (2009). Evaluating search results: An empirical analysis of middle school students' use of source attributes to select useful sources. *Journal of Educational Computing Research, 41*(1), 63–82. doi:10.2190/EC.41.1.c

Bråten, I., Strømsø, H.I., & Britt, M.A. (2009). Trust matters: Examining the role of source evaluation in students' construction of meaning within and across multiple texts. *Reading Research Quarterly, 44*(1), 6–28. doi:10.1598/RRQ.44.1.1

Brem, S.K., Russell, J., & Weems, L. (2001). Science on the Web: Evaluation of scientific arguments by a lay audience. *Discourse Processes, 32*(2), 191–213.

Britt, M.A., & Aglinskas, C. (2002). Improving students' ability to identify and use source information. *Cognition and Instruction, 20*(4), 485–522. doi:10.1207/S1532690XCI2004_2

Chafe, W.L. (1994). *Discourse, consciousness and time: The flow and displacement of conscious experience in speaking and writing.* Chicago: University of Chicago Press.

Chi, M.T.H. (1997). Quantifying qualitative analyses of verbal data: A practical guide. *The Journal of the Learning Sciences, 6*(3), 271–315. doi:10.1207/s15327809jls0603_1

Chi, M.T.H. (2000). Self-explaining: The dual processes of generating inferences and repairing mental models. In R. Glaser (Ed.), *Advances in instructional psychology: Vol. 5. Educational design and cognitive science* (pp. 161–238). Mahwah, NJ: Erlbaum.

Chi, M.T.H., De Leeuw, N., Chiu, M., & Lavancher, C. (1994). Eliciting self-explanations improves understanding. *Cognitive Science, 18*(3), 439–477.

Chinn, C.A., & Malhotra, B.A. (2002). Epistemologically authentic inquiry in schools: A theoretical framework for evaluating inquiry tasks. *Science Education, 86*(2), 175–218. doi:10.1002/sce.10001

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Erlbaum.

Coiro, J., & Dobler, E. (2007). Exploring the online reading comprehension strategies used by sixth-grade skilled readers to search for and locate information on the Internet. *Reading Research Quarterly, 42*(2), 214–257. doi:10.1598/RRQ.42.2.2

Coté, N., & Goldman, S.R. (1999). Building representations of informational text: Evidence from children's think-aloud protocols. In H. van Oostendorp & S.R. Goldman (Eds.), *The construction of mental representations during reading* (pp. 151–174). Mahwah, NJ: Erlbaum.

Coté, N., Goldman, S.R., & Saul, E.U. (1998). Students making sense of informational text: Relations between processing and representation. *Discourse Processes, 25*(1), 1–53. doi:10.1080/01638539809545019

Ferguson, L.E., Bråten, I., & Strømsø, H.I. (2012). Epistemic cognition when students read multiple documents containing conflicting scientific evidence: A think-aloud study. *Learning and Instruction, 22*(2), 103–120. doi:10.1016/j.learninstruc.2011.08.002

Gerjets, P., Kammerer, Y., & Werner, B. (2011). Measuring spontaneous and instructed evaluation processes during Web search: Integrating concurrent thinking-aloud protocols and eye-tracking data. *Learning and Instruction, 21*(2), 220–231. doi:10.1016/j.learninstruc.2010.02.005

Gobert, J.D. (2000). A typology of causal models for plate tectonics: Inferential power and barriers to understanding. *International Journal of Science Education, 22*(9), 937–977. doi:10.1080/095006900416857

Goldman, S.R. (2004). Cognitive aspects of constructing meaning through and across multiple texts. In N. Shuart-Faris & D. Bloome (Eds.), *Uses of intertextuality in classroom and educational research* (pp. 317–352). Greenwich, CT: Information Age.

Goldman, S.R. (with Lawless, K.A., Gomez, K.W., Braasch, J., MacLeod, S., & Manning, F.). (2010). Literacy in the digital world: Comprehending and learning from multiple sources. In M.G. McKeown & L. Kucan (Eds.), *Bringing reading research to life* (pp. 257–284). New York: Guilford.

Goldman, S.R., & Rakestraw, J.A., Jr. (2000). Structural aspects of constructing meaning from text. In M.L. Kamil, P.B. Mosenthal, P.D. Pearson, & R. Barr (Eds.), *Handbook of reading research* (Vol. 3, pp. 311–335). Mahwah, NJ: Erlbaum.

Goldman, S.R., & Saul, E.U. (1990). Flexibility in text processing: A strategy competition model. *Learning and Individual Differences, 2*(2), 181–219. doi:10.1016/1041-6080(90)90022-9

Graesser, A.C., Singer, M., & Trabasso, T. (1994). Constructing inferences during narrative text comprehension. *Psychological Review, 101*(3), 371–395. doi:10.1037/0033-295X.101.3.371

Graesser, A.C., Wiley, J., Goldman, S.R., O'Reilly, T., Jeon, M., & McDaniel, B. (2007). SEEK Web tutor: Fostering a critical stance while exploring the causes of volcanic eruption. *Metacognition and Learning, 2*(2/3), 89–105. doi:10.1007/s11409-007-9013-x

Griffin, T.D., Wiley, J., & Salas, C. (in press). Supporting effective self-regulated learning: The critical role of monitoring. In R. Azevedo & V. Aleven (Eds.), *International handbook of metacognition and learning technologies.* New York: Springer Science.

Griffin, T.D., Wiley, J., & Thiede, K.W. (2008). Individual differences, rereading, and self-explanation: Concurrent processing and cue validity as constraints on metacomprehension accuracy. *Memory & Cognition, 36*(1), 93–103. doi:10.3758/MC.36.1.93

Hartman, D.K. (1995). Eight readers reading: The intertextual links of proficient readers reading multiple passages. *Reading Research Quarterly, 30*(3), 520–561. doi:10.2307/747631

Hemmerich, J.A., & Wiley, J. (2002). Do argumentation tasks promote conceptual change about volcanoes? In W.D. Gray & C.D. Schunn (Eds.), *Proceedings of the twenty-fourth annual* *conference of the Cognitive Science Society* (pp. 453–458). Hillsdale, NJ: Erlbaum.

Jones, S. (2002). *The Internet goes to college: How students are living in the future with today's technology*. Washington, DC: Pew Internet & American Life Project. Retrieved July 18, 2012, from www.pewinternet.org/~/media//Files/Reports/2002/PIP_College_Report.pdf.pdf

Kiili, C., Laurinen, L., & Marttunen, M. (2008). Students evaluating Internet sources: From versatile evaluators to uncritical readers. *Journal of Educational Computing Research, 39*(1), 75–95. doi:10.2190/EC.39.1.e

Kintsch, W. (1988). The role of knowledge in discourse comprehension: A construction-integration model. *Psychological Review, 95*(2), 163–182. doi:10.1037/0033-295X.95.2.163

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

Kuiper, E., Volman, M., & Terwel, J. (2005). The Web as an information resource in K–12 education: Strategies for supporting students in searching and processing information. *Review of Educational Research, 75*(3), 285–328. doi:10.3102/00346543075003285

Lawless, K.A., Goldman, S.R., Gomez, K., Manning, F., & Braasch, J. (in press). Assessing multiple source comprehension through evidence centered design. In J.P. Sabatini, E. Albro, & T. O'Reilly (Eds.), *Reaching an understanding: Innovations in how we view reading assessment*. Lanham, MD: R&L Education.

Leinhardt, G., & Young, K.M. (1996). Two texts, three readers: Distance and expertise in reading history. *Cognition and Instruction, 14*(4), 441–486. doi:10.1207/s1532690xci1404_2

Lenhart, A., Rainie, L., & Lewis, O. (2001). *Teenage life online: The rise of the instant-message generation and the Internet's impact on friendships and family relationships*. Washington, DC: Pew Internet & American Life Project. Retrieved July 18, 2012, from www.pewinternet.org/~/media//Files/Reports/2001/PIP_Teens_Report.pdf.pdf

Leu, D.J., Coiro, J., Castek, J., Hartman, D.K., Henry, L.A., & Reinking, D. (2008). Research on instruction and assessment in the new literacies of online reading comprehension. In C.C. Block & S.R. Parris (Eds.), *Comprehension instruction: Research-based best practices* (2nd ed., pp. 321–346). New York: Guilford.

Magliano, J.P., & Millis, K.K. (2003). Assessing reading skill with a think-aloud procedure and latent semantic analysis. *Cognition and Instruction, 21*(3), 251–283. doi:10.1207/S1532690XCI2103_02

Marshak, S. (2001). *Earth: Portrait of a planet*. New York: W.W. Norton.

Mason, L., Ariasi, N., & Boldrin, A. (2011). Epistemic beliefs in action: Spontaneous reflections about knowledge and knowing during online information searching and their influence on learning. *Learning and Instruction, 21*(1), 137–151. doi:10.1016/j.learninstruc.2010.01.001

Mason, L., Boldrin, A., & Ariasi, N. (2010). Searching the Web to learn about a controversial topic: Are students epistemically active? *Instructional Science, 38*(6), 607–633. doi:10.1007/s11251-008-9089-y

McNamara, D.S. (2004). SERT: Self-explanation reading training. *Discourse Processes, 38*(1), 1–30. doi:10.1207/s15326950dp3801_1

McNamara, D.S., & Magliano, J.P. (2009). Self-explanation and metacognition: The dynamics of reading. In D.J. Hacker, J. Dunlosky, & A.C. Graesser (Eds.), *Handbook of metacognition in education* (pp. 60–81). New York: Routledge.

Moje, E.B. (2008). Foregrounding the disciplines in secondary literacy teaching and learning: A call for change. *Journal of Adolescent & Adult Literacy, 52*(2), 96–107. doi:10.1598/JAAL.52.2.1

Ozuru, Y., Briner, S., Best, R., & McNamara, D.S. (2010). Contributions of self-explanation to comprehension of high- and low-cohesion texts. *Discourse Processes, 47*(8), 641–667. doi:10.1080/01638531003628809

Perfetti, C.A., Rouet, J., & Britt, M.A. (1999). Toward a theory of documents representation. In H. van Oostendorp & S.R. Goldman (Eds.), *The construction of mental representations during reading* (pp. 88–108). Mahwah, NJ: Erlbaum.

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

Rouet, J., & Britt, M.A. (2011). Relevance processes in multiple document comprehension. In M.T. McCrudden, J.P. Magliano, & G. Schraw (Eds.), *Text relevance and learning from text* (pp. 19–52). Charlotte, NC: Information Age.

Rouet, J., Favart, M., Britt, M.A., & Perfetti, C.A. (1997). Studying and using multiple documents in history: Effects of discipline expertise. *Cognition and Instruction, 15*(1), 85–106. doi:10.1207/s1532690xci1501_3

Sanchez, C.A., Wiley, J., & Goldman, S.R. (2006). Teaching students to evaluate source reliability during Internet research tasks. In S.A. Barab, K.E. Hay, N.B. Songer, & D.T. Hickey (Eds.), *Making a Difference: Vol. 1. The proceedings of the seventh International Conference of the Learning Sciences* (pp. 662–666). New York: Routledge.

Shanahan, T., & Shanahan, C. (2008). Teaching disciplinary literacy to adolescents: Rethinking content-area literacy. *Harvard Educational Review, 78*(1), 40–59.

Stadtler, M., & Bromme, R. (2004). Laypersons searching for medical information on the Web: The role of metacognition. In K. Forbus, D. Gentner, & T. Regier (Eds.), *Proceedings of the twenty-sixth annual conference of the Cognitive Science Society* (p. 1638), Chicago: Erlbaum.

Stadtler, M., & Bromme, R. (2007). Dealing with multiple documents on the WWW: The role of metacognition in the formation of documents models. *International Journal of Computer-Supported Collaborative Learning, 2*(2/3), 191–210. doi:10.1007/s11412-007-9015-3

Stadtler, M., & Bromme, R. (2008). Effects of the metacognitive computer-tool *met.a.ware* on the Web search of laypersons. *Computers in Human Behavior, 24*(3), 716–737. doi:10.1016/j.chb.2007.01.023

Strømsø, H.I., Bråten, I., & Samuelstuen, M.S. (2003). Students' strategic use of multiple sources during expository text reading: A longitudinal think-aloud study. *Cognition and Instruction, 21*(2), 113–147. doi:10.1207/S1532690XCI2102_01

Thiede, K.W., Anderson, M.C.M., & Therriault, D. (2003). Accuracy of metacognitive monitoring affects learning of texts. *Journal of Educational Psychology, 95*(1), 66–73. doi:10.1037/0022-0663.95.1.66

Thiede, K.W., Redford, J.S., Wiley, J., & Griffin, T.D. (2012). Elementary school experience with comprehension testing may influence metacomprehension accuracy among seventh and eighth graders. *Journal of Educational Psychology*. Advance online publication. doi:10.1037/a0028660

Thomm, E., & Bromme, R. (2012). "It should at least seem scientific!" Textual features of "scientificness" and their impact on lay assessments of online information. *Science Education, 96*(2), 187–211. doi:10.1002/sce.20480

van den Broek, P., Risden, K., & Husebye-Hartmann, E. (1995). The role of readers' standards for coherence in the generation of inferences during reading. In R.F. Lorch, Jr., & E.J. O'Brien (Eds.), *Sources of coherence in reading* (pp. 353–373). Hillsdale, NJ: Erlbaum.

Wallace, R.M., Kupperman, J., Krajcik, J., & Soloway, E. (2000). Science on the Web: Students online in a sixth-grade classroom. *The Journal of the Learning Sciences, 9*(1), 75–104. doi:10.1207/s15327809jls0901_5

Walraven, A., Brand-Gruwel, S., & Boshuizen, H.P.A. (2009). How students evaluate information and sources when searching the World Wide Web for information. *Computers & Education, 52*(1), 234–246. doi:10.1016/j.compedu.2008.08.003

Wiley, J., Ash, I.K., Sanchez, C.A., & Jaeger, A. (2011). Clarifying goals of reading for understanding from expository science text. In M.T. McCrudden, J.P. Magliano, & G. Schraw (Eds.), *Text relevance and learning from text* (pp. 353–374). Charlotte, NC: Information Age.
Wiley, J., Goldman, S.R., Graesser, A.C., Sanchez, C.A., Ash, I.K., & Hemmerich, J.A. (2009). Source evaluation, comprehension, and learning in Internet science inquiry tasks. *American Educational Research Journal, 46*(4), 1060–1106. doi:10.3102/0002831209333183
Wiley, J., Griffin, T.D., & Thiede, K.W. (2005). Putting the comprehension in metacomprehension. *The Journal of General Psychology, 132*(4), 408–428. doi:10.3200/GENP.132.4.408-428
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*(2), 301–311. doi:10.1037/0022-0663.91.2.301
Wineburg, S.S. (1991). Historical problem solving: A study of the cognitive processes used in the evaluation of documentary and pictorial evidence. *Journal of Educational Psychology, 83*(1), 73–87. doi:10.1037/0022-0663.83.1.73
Wineburg, S. (1998). Reading Abraham Lincoln: An expert/expert study in the interpretation of historical texts. *Cognitive Science, 22*(3), 319–346. doi:10.1207/s15516709cog2203_3
Winne, P.H., & Hadwin, A.F. (1998). Studying as self-regulated learning. In D.J. Hacker, J. Dunlosky, & A.C. Graesser (Eds.), *Metacognition in educational theory and practice* (pp. 277–304). Mahwah, NJ: Erlbaum.
Wolfe, M.B.W., & Goldman, S.R. (2005). Relations between adolescents' text processing and reasoning. *Cognition and Instruction, 23*(4), 467–502. doi:10.1207/s1532690xci2304_2
Yore, L.D., Bisanz, G.L., & Hand, B.M. (2003). Examining the literacy component of science literacy: 25 years of language arts and science research. *International Journal of Science Education, 25*(6), 689–725. doi:10.1080/09500690305018

Submitted November 12, 2011
Final revision received June 4, 2012
Accepted June 15, 2012

Susan R. Goldman is a distinguished professor of liberal arts and sciences, psychology, and education at the University of Illinois at Chicago, USA; e-mail sgoldman@uic.edu.

Jason L.G. Braasch is a postdoctoral research fellow at the University of Oslo, Norway; e-mail jason.braasch@ped.uio.no.

Jennifer Wiley is a professor of psychology at the University of Illinois at Chicago, USA; e-mail jwiley@uic.edu.

Arthur C. Graesser is a professor in the Department of Psychology and the Institute of Intelligent Systems at the University of Memphis, Tennessee, USA; e-mail art.graesser@gmail.com.

Kamila Brodowinska is a doctoral candidate in the Educational Psychology Program, College of Education, University of Illinois at Chicago, USA; e-mail kbrodo2@uic.edu.

Appendix A

Think-Aloud Instruction

What you find out in your research is one thing we are interested in in this experiment. We are also interested in how you do your research—that includes what you decide to read and how you make those decisions. So you'll read aloud or say out loud what you are focusing on. And we want you to tell us anything the information makes you think while you are doing your research. This is called Thinking Aloud. We want you to Think Aloud while you are deciding what to read and when you are actually reading information on a website.

So when you look at the Google page or any of the website pages that you select, we want you to tell us what you are looking at and what you are thinking about that information. If you are scanning a page, we want you to tell us you are scanning ("I'm looking at the title of the website," or "I'm looking at the menu bar on the left of the website and see that it says…"). We also want you to tell us what you are thinking about the information you scan and why you are thinking that ("I'm thinking that this is relevant because…," or "This menu bar is just about how to contact the folks who put the page up").

When you are actually reading something you are going to read it aloud and comment on what you are reading. This will help us know exactly what you are focusing on. You should make comments whenever you wish but try to comment on each sentence you read or diagram that you look at. Your comments can be whatever comes to mind as you read the information. Things that come to mind might be other things you've read or seen, experiences you've had, how the information helps—or doesn't help—you prepare for your writing task, what you understand from the information, how you are understanding the information. So just about anything that comes to mind. To help you think aloud every once in awhile we'll ask you, "What are you thinking? What does the information make you think of?"

We'll try it with a sample now. Let's imagine you are writing a report on metabolism for a group of adolescents who are interested in weight control. Here is a Google page that resulted from putting the keywords "metabolism" and "weight control" in.

> Metabolism. [I don't know anything about metabolism.]
> Customers in many pharmacies may soon be seeing the latest in new devices for the health conscious. [So it has something to do with health.] A sports psychologist is developing a Tab O Meter, a device he hopes will measure the human body's ability to produce energy efficiently. [OK, so the new device is the metab-O-meter. I bought one of those

stomach exercisers, and it was a waste of money.] The rate at which the body produces energy is called metabolism. [OK, I think I remember reading about that in my health class.] Different people have different metabolic rates that indicate how easily they can produce energy. [That makes sense, like, because some people seem to have more energy. I guess if you're into sports, you want to have a lot of energy.]

[Experimenter scans down to bottom of the page.] There are several factors...Another factor— [OK, so I'm looking at the beginning of each paragraph to see what else is likely to be in here.] Another factor affecting metabolism is the climate of the environment. [tropical, arctic, dry, humid] Temperature may cause metabolism to change. [I bet if it gets hot, metabolism increases.] People and animals that live in cold environments need to produce more energy in order to keep warm. [In that case, if it gets hot, then the rate will decrease.] Most animals that live in polar regions have high metabolisms. [Right, that makes sense with the last sentence.] If people move from a warm climate to a cold climate [Maybe metabolic rate isn't affected by temporary change.], their metabolic rates will increase. [That's why large people sweat so easily…they have a high metabolism.] Metabolic rate also differs depending on activity level. [Up here it says that "different people have different metabolic rates that indicate how easily they can produce energy," so the more you sweat, the more energy you produce…I can see that.]

OK, now you try. Read each sentence out loud and then comment after each sentence.

Appendix B

Causal Model of the Eruption of Mt. St. Helens