Improving Students’ Ability to Identify and Use Source Information

M. Anne Britt and Cindy Aglinskas

Northern Illinois University

Sourcing, contextualization, and corroboration are document-level literacy skills that experts routinely use when working with history documents. These are also skills that educators and curricula planners expect students to acquire. We examine high school and college students’ current degree of proficiency with these skills and describe our development and evaluation of a computer-based tutoring system designed to teach these skills. We observed that high school and college students who were asked to read multiple documents did not spontaneously attend to source information. Based on analyses of expert and intermediate behavior, we developed the Sourcer’s Apprentice, a computer-based tutorial and practice environment for teaching students to source and corroborate while reading history texts. In 3 evaluation studies we found that students who used the Sourcer’s Apprentice in place of regular classroom activity or a textbook-centered version of the same content improved at sourcing, contextualization, and corroboration on a transfer test. The Sourcer’s Apprentice group also wrote essays on the topic that were more integrated, cited more sources, and referenced more information from primary and secondary sources than the comparison group.

Current U.S. national standards in history recommend that students at all grade levels learn to evaluate, interpret, and compare textual information as well as construct their own interpretations of material provided in class (National Center for History in the Schools, 1996). This clearly presents a challenge to educators, considering students’ current ability level and experience with applying such document-level skills. The most recent national assessment shows that only 10% of 12th graders were deemed at or above the proficient level of skills such as using historical evidence to support positions, producing written arguments that

Requests for reprints should be sent to M. Anne Britt, 363 Psychology-Math Building, Northern Illinois University, Dekalb, IL 60115. E-mail: britt@niu.edu
reflect an in-depth grasp of issues, and referring to sources (National Assessment of Educational Progress, 1996). Meeting this challenge is even more daunting given the low likelihood that extended class time and resources will be devoted to teaching document-level skills.

The focus of this article is on a set of important but often neglected document-level literacy skills that are critical to achieving our national history standards. In particular, we focus on helping students to notice features of a source, to use source information to evaluate content and place events in a correct temporal context, and finally to compare information across documents. We begin with a detailed analysis of the component skills required to use sources. Next we present an experiment that examines the extent to which high school and college students spontaneously attend to source information in a multiple-document learning task. Then we describe our efforts to improve students’ ability to attend to and use source information through a computer-based tutoring system we call the Sourcer’s Apprentice. This tutor is a rather simple computer application that was designed based on principles from cognitive-instructional learning theories. Finally, we conclude with results from three effectiveness studies that evaluated the impact of the Sourcer’s Apprentice in history classes.

**DOCUMENT-LEVEL READING SKILLS**

Studies comparing high school students with expert historians have shown that novices do not approach documents in the same manner as experts (Rouet, Favart, Britt, & Perfetti, 1997; Wineburg, 1991). Wineburg (1991) found that unlike high school students, historians sought out and evaluated the source of a document and used this source information to influence their interpretation of the document’s content (referred to by Wineburg as a **sourcing heuristic**). Historians also attempted to situate the events in an accurate spatial-temporal context (referred to by Wineburg as a **contextualization heuristic**). Finally, Wineburg noted that experts directly compared content across documents systematically and were therefore able to identify discrepancies (referred to by Wineburg as a **corroboration heuristic**). Novices failed to systematically use any of these heuristics when reading documents. For instance, novices failed to attend to source information, tended to view their textbook as more trustworthy than secondary or primary sources, and had difficulty resolving, and sometimes even acknowledging, discrepancies among sources (Perfetti, Britt, & Georgi, 1995; Perfetti, Britt, Rouet, Georgi, & Mason, 1994; Wineburg, 1991). Such expert strategies are precisely the type of document-level literacy skills called for in the national standards, but are clearly lacking even among advance placement American history seniors. We now explore the components that are
necessary to skillfully source, contextualize, and corroborate information from multiple texts.

Sourcing

Evaluation of source information depends upon initially identifying critical features of the source. Experts attend to many features of sources, and some, such as style, may be too subtle for high school students to focus on early in their learning. In an effort to determine which elements of a source would be most helpful for students to attend to, we reanalyzed data from a previous multiple-document reading study (Rouet, Britt, Mason, & Perfetti, 1996). Rouet and colleagues required undergraduate students to read a set of related documents and then, from memory, rank the documents in terms of their usefulness and trustworthiness and justify each ranking. Of the nearly 1,500 justifications produced, 24% mentioned aspects of the author and 14% mentioned characteristics of the document. The remaining justifications were focused on content, task, and subject. We categorized the justifications regarding author and document to identify the type of source characteristics that students attended to most frequently. Definitions of each category and examples from the undergraduates’ statements can be found in Table 1. We expect that these characteristics are only a subset of those attended to by experts, but these are probably the set that could be most easily learned by novices.

Many of the justifications explicitly mentioned or evaluated the author of the document, including statements about the author’s position, motivation, or participation, or an evaluation of the author in general. **Position** is a reference to an author’s occupation, profession, or credentials. This is important because it indicates the author’s educational training, affiliation with an organization or institution, and status as an authority. **Motivation** is a possible reason why the author may have written the document. Author motives are usually not explicitly mentioned and therefore the reader must infer them from knowledge about the author or the type of document. Identifying possible motives is important for explaining why specific events or facts are mentioned whereas others are omitted, as well as for explaining why different persons and events are portrayed as they are. **Participation** is an indication that the student knew that the author was (or was not) a participant in the activities described. This category may be expanded to include a more general reference to how the author came to know the information described, such as direct experience, scholarly research, newspaper investigation, or fictional invention. Finally, **author evaluation** is a statement in which students state their own opinion of the author.

Many of the justifications referred not to author characteristics, but to characteristics of the document. The most frequently mentioned characteristics of the
document were date and document type. General document evaluations were also common. Date is an explicit mention of the time period in which the document was written. Date is important because it helps to indicate how much information the writer could access and how accurate an account of the events from memory might be. Also, the author’s perspective might be influenced by the time period in which the document was written. Document type is a reference to the form of the document, such as textbook, autobiography, treaty, scholarly book, or sworn deposition. The type of document can provide information about whether it is a record without interpretation, interpretation evaluated by informed peers, a concise overview or

<table>
<thead>
<tr>
<th>Feature</th>
<th>Definition</th>
<th>Examples</th>
</tr>
</thead>
<tbody>
<tr>
<td>Position</td>
<td>Occupation or credentials</td>
<td>• President</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Professor</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Railroad man</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Senator</td>
</tr>
<tr>
<td>Motivation</td>
<td>Reasons for the author writing the document</td>
<td>• Wanted to make himself look good</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Very biased written [sic] in his best interest to cover his own ass</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• President would say anything to keep from being impeached</td>
</tr>
<tr>
<td>Participation</td>
<td>How author came to know about events that are described</td>
<td>• Bystander</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Eye-witness account</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• He doesn’t exactly know what happened, receiving hearsay</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• He had firsthand information</td>
</tr>
<tr>
<td>Author evaluation</td>
<td>Critique or opinion of the author</td>
<td>• Blind fool, accepts lies</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Author is biased</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Just doesn’t want to believe our government</td>
</tr>
<tr>
<td></td>
<td></td>
<td>is corrupt</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Too schizophrenically written to be a lie, he is paranoid</td>
</tr>
<tr>
<td>Date</td>
<td>Time period in which document was written</td>
<td>• Author is writing after occurrence and has liberty to pick and choose</td>
</tr>
<tr>
<td></td>
<td></td>
<td>what he wishes to include in forming his opinion</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Different time period, how does he know</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Written with the ability to see retrospectively</td>
</tr>
<tr>
<td>Document type</td>
<td>Form of document</td>
<td>• Personal letter</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Official record</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Actual treaty</td>
</tr>
<tr>
<td>Document evaluation</td>
<td>Critique or opinion of type or date</td>
<td>• College text would not print false facts</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Official records make the most unbiased sources</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• As a treaty it has no bias</td>
</tr>
</tbody>
</table>
summary of agreed-upon information, or mere entertainment. Finally, a document evaluation is a statement of the student’s opinion of the document type.

Undoubtedly there are additional, more subtle, source characteristics to which experts attend, such as linguistic style, cultural setting, and intended audience. However, this set of characteristics provides a glimpse of the source features that are salient to students, and thus provides a possible starting point for instruction.

Contextualization

Contextualization is the heuristic of attending to a detailed sequence of events in time and space when reconstructing the events from documents. The reader attempts to create an elaborate model of the described scenario by incrementally analyzing the document and simulating the stated events to test hypotheses of plausible scenarios and reject other models as implausible. This requires the reader to go beyond the actual words stated in the document and therefore requires extensive knowledge and experience. For example, Wineburg (1994) documented one historian considering the logistics of moving 1,000 troops through a swamp in the middle of winter in 18th-century New England, as well as the probable duration of such an event and the likely physical and emotional effects on the soldiers. From an analysis of two experts’ think-aloud protocols, Wineburg (1998) identified six categories of statements that reflect the contextualization heuristic. These categories include spatiotemporal, social–rhetorical, biographic, historiographic, linguistic, and analogical comments. He also found that although a great deal of factual knowledge may be required to use such categories to establish context (e.g., biographical information), a critical element of expertise appears to be the ability to identify holes in the expert’s own knowledge, formulate questions, and mark information or conclusions as questionable or tentative.

Because situating events in an accurate spatiotemporal context requires the reader to elaborate document content with detailed and often subtle knowledge that most high school students lack (Wineburg, 1991), our focus on this heuristic is limited to learning to identify relevant source features that can later be used to create an appropriate context. In fact, Wineburg and Fournier (1994) found that even teachers-in-training either failed to create a context for historical events or selected a contemporary context. We therefore considered conceptualization, as discussed by Wineburg (1994, 1998), to be a more complex skill than sourcing. For purposes of the present study, we broadened the conceptualization of sourcing and focus on the skill of contextualization only as it relates to identifying relevant features of a source that can be useful in creating a context for historical information. Our broader definition of sourcing starts with Wineburg’s (1991) definition, “When evaluating historical documents, look first
to the source or attribution of the document” (p. 79), and expands it to include any aspect of the process of identifying, evaluating, and using information about a source (i.e., the author as well as the document itself). In defining sourcing in this way, we are including much information that would be useful or necessary to the application of the contextualization heuristic. For example, source features such as a document’s date could be used to construct an accurate event time line as well as to interpret the particular linguistic context of phrases based on their usage in the period when written. Similarly, focusing on author motivations could help the reader create a social–rhetorical context to evaluate and interpret content.

Corroboration

Corroboration is the general skill of checking facts or interpretations from a particular document against other, independent sources. Corroboration involves directly comparing the information from the various sources to identify which important statements are agreed on, which are uniquely mentioned, and which are discrepant. It enables readers to differentially weight the credibility of the information presented in a document. One’s confidence in a document’s credibility can be bolstered by the accumulation of corroborated information (i.e., information in agreement with other texts). Thus, when students find a fact that another author mentioned, they can have more confidence in the accuracy of that information.

Often, however, information is not corroborated, and as a result the student may judge that information to be tentative until corroborating information is located. There are three types of uncorroborated information: unique, contradicted, and incomplete. First, a document may present unique information not presented by other documents. No single document will provide all possible details and causal explanations for the described situation. Unique information is helpful because it allows the student to fill in a more complete set of events and motives. Unique information, however, also places a burden on the reader to hold that information as tentative until corroboration has been obtained. Second, a document may present information that contradicts information from other sources. The process of corroborating information against other texts allows the reader to identify such discrepancies. In this case, rather than merely marking the information as tentative, the reader must resolve the discrepancies and attempt to investigate the cause of the discrepancies to determine the extent to which an author is being deceptive or purposefully misleading. Finally, a document may present an incomplete version of the story by omitting important, generally agreed upon facts. This may lead the reader to question the credibility of the source in general. In the next section we consider students’ ability to identify and use source information in a multiple-document situation.
How Effectively Do High School Students Use Source Information?

Wineburg’s (1991) study suggested that high school students do not spontaneously apply sourcing, contextualization, and corroboration when reading documents. The advanced placement high school students in Wineburg’s study did not seek out source information prior to reading, rated the textbook as more trustworthy than primary documents, and based their criteria for evaluation more on content than on features of the source. Furthermore, they did not seek out corroborative information prior to incorporating new information into their representation of the story, nor did they note critical discrepancies among documents.

Rouet et al. (1996), however, found evidence that college students do have some ability to use sourcing strategies when they are provided with a supportive environment. Rouet et al. found that students who had access to primary documents rated them as more useful and trustworthy than other document types (e.g., historical essays and textbooks) and their justifications of their rankings focused on essential features of the author (e.g., author’s credentials, motivations, or participation in the events) and document type (e.g., when it was written). Compared with students who did not read primary documents, students who did included more citations in their essays, and most of the citations were from the primary documents. Most (65%) of the essays written, from memory, by the primary-source group included references to other documents, and most of the citations were accurate, indicating that the college students were attending to sources during their analysis of the documents. Additional evidence that novices have at least limited ability to source comes from Kushner (1996), who found that both undergraduate and graduate students spontaneously referred to characteristics of a document’s author during a think-aloud reading.

The apparent conflict between the Wineburg (1991) and the Rouet et al. (1996) findings may be partly due to differences in environmental support for applying document-level strategies. Rouet et al. did several things that may have supported the application of these strategies. For example, the documents given to students were all somewhat relevant and trustworthy (i.e., no historical fiction, as in Wineburg, 1991), and the texts were ordered from most general to most specific. Access to the documents also supported sourcing because students were forced to view the extended source before reading the document, whereas the sources for the Wineburg study were presented at the end of each document. Rouet et al. (1996) also used brief source citations to organize the documents, perhaps focusing the students on author and document characteristics. Finally, Rouet et al. organized the document set around a clearly defined controversy, which may have signaled the importance of source evaluation. Thus, it is unclear whether undergraduates would have been as proficient at sourcing had they not had such environmental supports available to them.
EXPERIMENT 1

In Experiment 1, we directly examined the extent to which high school and college students attend to source information when support is not available. As in the Wineburg (1991) study, source information was presented at the end of each excerpt, not all documents were trustworthy, and the documents were randomly ordered. All of the high school students and half of the college students were given comprehension instructions (i.e., read the excerpts to understand what happened and why), whereas the other half of the college students were given sourcing instructions (i.e., attend to information about the authors of the excerpts and take into account their bias or lack of knowledge while learning what happened and why). The comprehension condition provides a baseline measure of the degree to which participants spontaneously represent and evaluate source information when reading multiple related texts. The sourcing condition enables us to find out how much students can improve if given only a simple exhortation to attend to sources. Finally, the inclusion of both high school and college students allowed us to examine whether development may account for differences between the Wineburg (1991) and the Rouet et al. (1996) studies.

Method

Participants. Sixty students from four 11th-grade American history classes from a small, upper middle class, suburban school outside Boston participated. The students’ teachers occasionally enhanced the textbook with the use of documents; however, the focus of the history instruction was predominately the textbook, and there was no use of multiple documents on the same topic. Students were tested in the middle of the school year, early March, as students were beginning their study of the early 20th century. The college group consisted of 49 undergraduate psychology majors from a research methods class at Northern Illinois University. All college students participated for course credit.

Materials. The topic of the document set was the events that led to the United States building a canal in Panama. The first page of the booklet provided instructions and two paragraphs of background information. Following pages included excerpts from six authentic texts\(^1\) that included historian analyses, participant accounts, and a novel. Documents were selected to ensure that there was a noticeable range on all the

\(^1\)President Theodore Roosevelt’s October 10, 1903, private letter to a friend (reprinted in Morison, 1951, p. 628); Professor William LaFeber’s 1978 book; Raymond Saunders’ 1995 novel; Professor Graham Stuart’s 1944 book; engineer-businessman Philippe Bunau-Varilla’s 1914 autobiography; President Roosevelt’s letter to a constituent on January 6, 1904 (reprinted in Morison, 1951, p. 689). See reference list for full citations.
various source features. For instance, we varied when the documents were written (i.e., at the time of the event vs. more recently) and the type of document (i.e., letters, historian essays, an autobiography, and a novel). The documents also varied with respect to their credibility. For example, authors included the U.S. president at the time, historians, and a novelist. The novelist clearly had no obligation to historical accuracy, and controversy exists about President Theodore Roosevelt’s intentions. Each excerpt was presented on a separate page and information about the source was included at the bottom of each page, thereby requiring the student to make a conscious decision to attend to the source information prior to reading. A single sheet of paper for note taking was provided with the packet.

A sourcing skills test was developed to assess our participants’ sourcing ability. The 14 source questions on this test ranged from noticing key elements of the source to directly evaluating the source. Examples of source identification questions include “Which document was written earliest?,” “Which author(s) were participants in the events in Panama?,” and “From what type of document was Raymond Saunders’ excerpt taken?” Examples of source evaluation include “Which document was the least trustworthy?” and “Which author(s) claimed that the United States helped plan or incite the Panamanian Revolution?” In addition to the source questions, the sourcing skills test also included a short essay question about the central narrative (“What happened in 1903 that enabled the United States to get a canal in Panama?”) and an essay question about the controversy (“To what extent were Roosevelt and his administration responsible for the 1903 revolution in Panama?”). Finally, the test included two recall questions that focused on the arguments underlying the controversy: “List 3 facts from the readings that could be used to show that the United States was responsible for inciting the revolt” and “List 3 facts from the readings that could be used to show that the United States was NOT responsible for inciting the revolt.”

**Design.** The high-school students ($n = 60$) and half of the undergraduates ($n = 24$) were given comprehension instructions. They were told to read for general understanding so they would be able to answer several questions about what happened and write an essay to answer each of the following questions: “What happened during the planning of the Panamanian revolution?” and “To what extent did the U.S. Government influence the planning of the revolution?” The remaining undergraduates ($n = 25$) were given sourcing instructions, which mentioned that each document provided a different perspective on the controversy and that they should pay attention to and evaluate the source of each document. They were told that they would have to answer questions about different authors’ perspectives as well as write two short essays.

**Procedure.** Students were given a booklet with six documents on separate pages and a single sheet on which they could take notes to use when answering
questions. The instructions, either comprehension or sourcing, were presented on the first page of the booklet. Both groups were told that their study time and note space were limited so they would have to be selective. After students finished reading, they handed in their text booklet but kept their note sheets. The experimenter then gave students a question booklet that contained the sourcing questions and the two essay questions. Participants were given 35 min to complete the entire task.

Results and Discussion

Sourcing scores were obtained by combining correct source information mentioned in note sheets (out of a possible 42 points) and the number of correct answers on the sourcing test (out of a possible 23 points). Table 2 shows the groups’ average scores out of a possible 65 points. Sourcing scores were fairly low for both high school and college students. The high school students who were given comprehension instructions mentioned only an average of 11.9% of the available information, whereas college students mentioned a slightly higher 16.2% when given comprehension instructions and 18.7% when given sourcing instructions. A between-subject ANOVA revealed a significant main effect of condition, $F(2, 106) = 6.82, p < .01$. A follow-up Neumann–Keuls test showed that high school students mentioned significantly less sourcing information than either group of college students. Within the college students, however, there was no difference among instructional conditions. Thus, students’ spontaneous use of source information is low and they require more than simple instructions to remedy this sourcing skill deficit.

A closer analysis of these sourcing scores reveals that students’ sourcing difficulty was reflected in both their note taking and their question answering. High school and college students included few source features in their notes, mentioning an average of only 10% and 14% of the 42 source features, respectively. They also failed to correctly answer a significant proportion of the source questions, averaging only 15% and 23% correct, respectively, on the 23-question test. Thus,

| TABLE 2 |
|-----------------|--------|--------|--------|--------|
| High school comprehension | 7.72   | 4.54   | 8.0    | 0–20   |
| Undergraduate comprehension | 10.56 | 5.18   | 10.0   | 2–22   |
| Undergraduate sourcing     | 12.13  | 6.94   | 11.0   | 1–26   |

*Out of total possible 65 points.*
neither group did a very good job of noting source information for later use or of correctly answering source questions.

Further evidence that students are not skilled sourcers comes from their answers regarding the controversy. The sourcing skills test included two argument questions that asked participants to list facts from the readings to support each position on the controversy. Recall that one of the documents in the set was a novel and was clearly described as such. Both 11th graders and college students provided information mentioned uniquely in the novel or made an explicit reference to novelist Saunders when giving facts that could be used to support an argument. Of the 38 high school and 41 college students answering at least one of these two argument questions, 32% of the high school students and 41% of the undergraduates answered with information that was unique to the novel. Inappropriate use of information was not restricted to answers to the argument questions. Consider the following excerpt from an essay written by one of the high school students:

Roosevelt and his administration were fairly responsible for these events. In Roosevelt’s letter to Dr. Shaw he says he would be delighted if Panama revolted, but that he couldn’t say it to the public. Roosevelt never verbally instigated the revolt, but he did so in other ways. In Raymond Saunders’ excerpt Bunau-Varilla told Roosevelt the isthmus would revolt if the United States backed them up with gun boats. Roosevelt implied the United States would, “all but winking” at him.

On the surface, this student appears to do an excellent job of citing a source to support a position. He uses an explicit reference, accurately paraphrases information from the cited document, and puts in quotes material taken verbatim from the source. The only problem is that the cited author, Raymond Saunders, is a novelist and the excerpt that the student repeated as fact was taken from a novel. Had the student attempted to find other authors who confirmed the accuracy of this information (i.e., corroboration), he would have realized that no other author mentioned this event. Thus, even students who are competent at citing a source may still not fully appreciate the document and author features that go into the selection of a source and the need to corroborate the information they use in their essays.

Although college students demonstrated greater sourcing skill than high school students, neither group performed at a level one would consider ideal. The top performer answered only 70% of the source questions correctly, indicating that both 11th graders and undergraduates would benefit from instruction in the use of sources. Furthermore, effective instruction would have to involve more than merely providing goals and a general orientation to sources.

In the following section we describe the Sourcer’s Apprentice, a computer-based tutor and practice environment designed to teach document-level skills by
aiding students in using source information when reading historical documents. We begin with a brief description of the instructional design principles that guided the design of the Sourcer’s Apprentice, followed by a description of its environment and instructional modules. Finally, we present several experiments that examine the effectiveness of the application.

THE SOURCER’S APPRENTICE: A COMPUTER-BASED ENVIRONMENT FOR TEACHING SOURCING SKILLS

The Sourcer’s Apprentice is a computer application that teaches sourcing, contextualization, and corroboration in the context of researching a historical controversy. Students read a set of highly structured and varied documents with the goal of writing an informed essay about the controversy. Although the Sourcer’s Apprentice is quite simple for a tutoring system, it nevertheless embodies many instructional principles from cognitive and educational theories. We summarize these later. For a more detailed discussion see Britt, Perfetti, Van Dyke, and Gabrys (2000).

Sourcer’s Apprentice Design Principles

_Principle 1: Teach through situated problem solving._ The Sourcer’s Apprentice takes its inspiration from the cognitive apprenticeship approach to learning. This approach advocates situating problem-solving activity in authentic contexts while using supports such as modeling, coaching, and fading to ease students into fully skilled performance (Brown, Collins, & Duguid, 1989; Collins, Brown, & Newman, 1989). The application models skilled sourcing by prompting students to fill in structured note cards with relevant source information and by including historian essays that act as models of how to interpret and cite other documents. Coaching is provided through student-requested help screens for each source feature and through incremental hints that are displayed following incorrect answers to questions. Finally, fading, though not presently part of the Sourcer’s Apprentice, could be accomplished by making all feature entries use typing, omitting interpretation of available primary documents, and requiring the students to use _Contemporary Authors_ or other authentic sources to obtain author and document information.

_Principle 2: Support expert representations._ Subject matter experts have highly organized and interconnected knowledge representations (Chi, Glaser, & Rees, 1982; Larkin, 1983; Lesgold, Feltovich, Glaser, & Wang, 1981). With respect to sourcing, expert representations have been characterized in terms such as a documents model (Britt, Perfetti, Sandak, & Rouet, 1999; Perfetti, Rouet,
Principle 3: Decompose the task. We based our selection of document and author features on our analysis of those used by undergraduates to justify their evaluations of documents (Rouet et al., 1996) and on Wineburg’s (1991) descriptions of how experts approached reading historical documents. By decomposing the skills and reflecting these elements prominently in the interface, Sourcer’s Apprentice scaffolds the student in seeing the importance of every component and focusing resources appropriately during learning.

Principle 4: Support transfer. Transfer is the ability to generalize the learned skill and apply it in an appropriate context. Although transfer is critical to producing flexible knowledge that can be applied in various real world situations, researchers have typically found it difficult to achieve (Gick & Holyoak, 1980; Hayes & Simon, 1977). The Sourcer’s Apprentice supports transfer in two complementary ways. First, the Sourcer’s Apprentice supports transfer through a mapping between the environment in which a skill is learned and the environment in which the skill must be applied (Singley & Anderson, 1989). The interface of the Sourcer’s Apprentice graphically and functionally resembles the situation that will most frequently demand the skill: the library setting with books and note cards. Second, the Sourcer’s Apprentice supports transfer through exposure to highly variable problem and text types to encourage the student to abstract highly general heuristics and concepts (Druckman & Bjork, 1991). Students receive different types of controversies (e.g., military, economic, and social history) and different types of documents (e.g., treaties, letters, congressional reports, historian essays, and political speeches). We expect that variation will help to foster a richly developed and flexible sense of source that is relatively independent of context-specific information.

Principle 5: Provide explicit instruction. One of the most effective ways to teach is through direct and explicit instruction. The Sourcer’s Apprentice tutorial explicitly informs students about source features and expert heuristics. Our approach is to begin with mandatory, minimalist instruction—just enough to get the student started and ensure that all students learn the basics.

Principle 6: Motivate engagement. There are several ways that the Sourcer’s Apprentice motivates students. The most obvious is the game environment, in which
students score points for activities that enhance the application of expert heuristics. We believe that it is critical that interface “bells and whistles” be used as reinforcers for successful application of newly learned skills rather than as window dressing or, even worse, distraction. Other motivation techniques identified by Malone and Lepper (1987) have also been incorporated into the Sourcer’s Apprentice. We present students with challenging goals that have uncertain outcomes, and we try to ensure ultimate success on the task by providing feedback and coaching so that the challenging goals do not lead to frustration and dropout. Students’ responses are scored immediately, enabling students to take corrective action if the response is incorrect (Anderson, Boyle, & Reiser, 1985; Lewis & Anderson, 1985), and feedback is sufficiently frequent and fine-grained to provide many small rewards for continuing. Finally, the Sourcer’s Apprentice is a mixed-initiative application, leaving students free to select their own course of investigation.

Sourcer’s Apprentice Environment

The basic idea behind the Sourcer’s Apprentice is to get students to read a structured set of documents about a single problem, have them identify source and document features along with important content, and then have them use this information to answer questions and write a short essay. These activities take place on a computer screen in the context of preparing a research paper in a “library” setting. As shown in Figure 1, the main screen of the Sourcer’s Apprentice resembles a library with a central bookshelf for selecting books and a set of individual note cards for taking notes.

Students using the Sourcer’s Apprentice for the first time begin with a tutorial that provides direct instruction on sourcing, contextualization, and corroboration. A single page of instruction is given for each feature, directing the student in three component skills for each feature: identification, use, and evaluation. Students must correctly answer two questions pertaining to the presented material before continuing. Thus, although the amount of initial instruction is limited, we ensure that it is understood before students proceed. We make additional, more detailed help available at any time from a scrollable pop-up help screen.

After the tutorial, the student sees the main screen, like the screen shown in Figure 1. The problem statement is always present in the top portion of the screen. Prominently displayed in the center of the screen is a “bookshelf” containing the books ordered from most general to most specific. To open a book, the student clicks on its image on the bookshelf and the book opens in the center of the screen (see Figure 2). Each “book” has four scrollable pages in the following order: title page, author page, document page, and content page. The author page provides detailed source information about the author’s credentials and possible motives, as shown in Figure 2. The document page explains the type of document, who published it, and
when it was written. The type of information presented on the author and document pages is information that one might obtain from the book’s dust jacket or reference sources; our pages were modeled after excerpts from *Contemporary Authors* with elaborations to allow the student to answer a set of questions about the source features. The content page displays the content of the excerpt. A student can change pages by clicking on the tabs at the bottom of the book. To support document comparison, the Sourcer’s Apprentice allows students to have two documents open on the screen simultaneously.

The problems and documents presented to students are carefully selected. Problems are historical controversies, that is, important historical events for which historians offer conflicting interpretations. Documents are excerpts from actual texts, chosen to ensure a variety of document types and arguments. Together the problem and documents make up a module. The present module set includes problems from military history (e.g., “To what extent were Roosevelt and his administration responsible for the 1903 revolution in Panama?”), economic history (e.g., “To what extent was Carnegie responsible for breaking the union at Homestead?”), and social history (e.g., “What was the primary cause of the Salem Witch Trials?”). Each of the existing modules contains seven
documents that vary in type. The first document is always a textbook excerpt that provides an overview of the situation, characters, and conflict. The next two documents are historian’s interpretations of the controversial event. The historian interpretations are selected to provide opposing accounts of the events, and they cite primary documents to support their arguments. Finally there are four primary documents that can be used as evidence to support either side of the controversy. Two of the primary documents are specifically mentioned in the historian’s accounts so the student has a model for citing and using primary documents as evidence.

The student’s task is to read each document to learn what it contributes to understanding the controversy. While reading, students fill in a note card for each book. Note cards appear at the bottom of the screen. Students can move among the cards, which are structured to ensure that students attend to source information as well as content. Each card contains a line for inserting information about six source and three content features. The source features are based on the categories in Table 1 and are grouped on the screen into three categories: author (who, position, how know, and author motives); document (when, type); and content (documents mentioned, main point, comments).
Information is inserted into a note card either by typing or by selecting a phrase from a page in the book and dropping it in a “bucket” icon. The example in Figure 2 shows the filling of the how know source feature. If the correct phrase is dropped into the bucket, an abstracted form of the information is inserted into the appropriate field of that book’s note card, and points are awarded to the student. When an incorrect answer is dropped, a different window pops up, deducting points and providing a hint toward the correct answer. Hints are presented in a graded fashion, from vague to specific, for each bucket. The first hint merely asks the student to try again. Each subsequent error results in a more specific hint until the correct answer is finally provided. All of the source features are filled using this drag-and-drop method except for why, main point, and comments. For these features, students directly type in their answers.

When the student decides he or she has studied the documents enough to form an educated opinion on the controversy, the student presses the done button. The student then receives a series of questions to answer about the sources and contents of the documents. Then the student receives the final essay environment in which the bookshelf is replaced by a large screen containing a field for the student to enter an essay. During the essay-writing phase, only the note cards are available to the student.

How well does the Sourcer’s Apprentice work? In Experiments 2A and 2B we compared student’s sourcing skill on a transfer test after either working with the Sourcer’s Apprentice or participating in regular classroom activity. The procedure and design were the same across Experiments 2A and 2B, with changes only in the school and the instructional module (i.e., topic). In Experiment 3 we examined the importance of the Sourcer’s Apprentice environment per se by comparing students who used the Sourcer’s Apprentice with another group of students who studied the same content structured as a textbook.

EXPERIMENT 2A

In this experiment an experimental group received a 2-day exposure to the Sourcer’s Apprentice and a control group received 2 days of regular classroom activities on the module topic. Both groups were given a sourcing test before and after the instructional session.

Method

Participants. The participants were students from two 11th-grade American history classes with one of the teachers from Experiment 1. Students were assigned to condition as intact classes. The 8 participants from the experimental group and
7 from the control group completed the entire 4 days of the experiment late in the school year.

**Module.** The Sourcer’s Apprentice module centered around the United States–Vietnam War. It consisted of seven authentic texts including excerpts from participant accounts, historian essays, and government reports helpful for answering the narrative question, “What happened at the Gulf of Tonkin during the summer of 1964 and why was it important?,” and the controversy, “What was the Gulf of Tonkin resolution and why did Lyndon B. Johnson push Congress to pass it in August of 1964?”

**Transfer tests.** Two 35-min tests were constructed around historical events unrelated to the topic of the training module (i.e., the Vietnam War). The topic of the pretest was the battle of Lexington–Concord and the topic of the posttest was the United States obtaining the rights to build a canal in Panama. The format of the transfer tests (both pretest and posttest) was identical to the sourcing skills assessment test described in Experiment 1. Students read six texts centered around a controversy while taking notes. After reading, students were asked several specific questions about the sources of the documents. Our definition of learning as the student’s ability to transfer to a novel problem in a different and unsupported context (e.g., without the structured interface) thus provides a rather conservative test of skill acquisition.

**Procedure.** On the first day all students, regardless of condition, were given the Lexington–Concord pretest. Students were given 35 min during a single class period to complete the pretest. During the next two 40-min class periods, both groups learned about Vietnam and the Gulf of Tonkin. The teacher routinely covered the

---

2Excerpts include Dr. Edwin E. Moise’s (1996) history book; Dr. George C. Herring’s (1986) history book; President Lyndon B. Johnson’s message to Congress on August 5, 1964 (reprinted in Galloway, 1970); Secretary of Defense Robert S. McNamara’s March 16, 1964, memorandum to President Johnson (reprinted in Sheehan, 1971); the Senate Foreign Relations Committee’s August 5, 1964, proposed congressional resolution (reprinted in Appleman, Williams, McCormick, Gardner, & LaFeber, 1985); the draft of a congressional resolution on May 25, 1964, by Assistant Secretary of State for East Asia and Pacific Affairs William Bundy (reprinted in Sheehan, 1971); and BBC radio 1977 interviews with Undersecretary of State George W. Ball and Assistant Secretary of State for East Asia and Pacific Affairs William Bundy (reprinted in Charlton & Moncrieff, 1978).

3For the Lexington–Concord test the narrative question was “What happened on the Lexington Green on April 19, 1775?” and the controversy question was “To what extent were the British responsible for the events of April 19, 1775?” This controversy and document set were adapted from Wineburg (1991). The six excerpts were from sworn testimony by 34 Minutemen on April 25, 1775 (Sawtell, 1968); Professor George Bancroft’s (reprinted from Bennet, 1970) book; British regular John Bateman’s sworn statement on April 23, 1775 (reprinted in Bennet, 1970); American History teacher Louis Birnbaum’s (1986) book; Howard Fast’s (1961) novel; and British Lieutenant William Sutherland’s April 27, 1775, letter to the secretary of the British commander (Bennet, 1970).
topic, and the control students stayed in their classroom and engaged in regular classroom activities. For the experimental students, the teacher provided background material in a lecture for 10 min at the beginning of the class and then students read a short background text written by the experimenters. Then students went through a brief, 20-min tutorial providing training on the environment and the basic elements of sourcing. Most students had only just begun working on the actual controversy when class ended. This left one full class period for students to work on the Sourcer's Apprentice module. All experimental students worked individually on a computer in a room attached to the school’s library. They read the seven documents and filled in the note cards for each document. Then they answered several content and source questions. Finally, they were given access to their note cards and asked to write an essay on the controversy. On the final day, after an intervening weekend, all students in both conditions completed the Panama Canal posttest in their regular classroom. No statements were made connecting the posttest to the computer activity.

Results and Discussion

A comparison of sourcing performance on the pretest and posttest was used to test the effectiveness of the Sourcer’s Apprentice. Scores for these tests were computed by adding the number of source features mentioned in the students’ notes for each document (e.g., credentials, date written, and type of document) to the number of source questions answered correctly. The means for each condition (Sourcer’s Apprentice vs. control) and test occasion (pretest vs. posttest) are shown in Table 3. A 2 (Condition) × 2 (Test Occasion) mixed design analysis of

<table>
<thead>
<tr>
<th>Test Occasion</th>
<th>Pretest</th>
<th>Posttest</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>M</td>
<td>SD</td>
</tr>
<tr>
<td>Experiment 2A</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sourcer’s Apprentice</td>
<td>14.38</td>
<td>6.35</td>
</tr>
<tr>
<td>Control group</td>
<td>16.57</td>
<td>7.14</td>
</tr>
<tr>
<td>Experiment 2B</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sourcer’s Apprentice</td>
<td>13.20</td>
<td>7.42</td>
</tr>
<tr>
<td>Control group</td>
<td>15.63</td>
<td>7.84</td>
</tr>
<tr>
<td>Experiment 3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sourcer’s Apprentice</td>
<td>8.22</td>
<td>6.42</td>
</tr>
<tr>
<td>Control group</td>
<td>9.79</td>
<td>4.02</td>
</tr>
</tbody>
</table>
variance (ANOVA) revealed no main effect of condition, $F(1, 13) = 1.85, p = .20$, and a marginally significant main effect of test occasion, $F(1, 13) = 3.99, p = .07$. The trend was toward more correct identification of source information in the posttest occasion ($M = 19.6$) than in the pretest occasion ($M = 15.4$). More importantly, there was a significant interaction of condition and test occasion, $F(1, 13) = 11.54, p < .01$. The Sourcer's Apprentice group correctly identified an average of 10.3 more items on the posttest compared to the pretest, whereas the control group correctly identified an average of 2.7 fewer items on the posttest. The decrease for the control group on the posttest is likely due to the students being more familiar with the battle of Lexington–Concord than the events surrounding the building of the Panama Canal. The relative difference between the two groups was 13 items or 19% more correct information for the Sourcer's Apprentice group. Thus, exposure to the Source’s Apprentice resulted in significantly improved sourcing on the transfer test.

**EXPERIMENT 2B**

Experiment 2B replicates Experiment 2A with a different population and a different instructional module.

**Method**

**Participants.** Two 11th-grade economics classes participated in early May. Both classes had the same teacher and were from a school in a small rural school district outside of Pittsburgh. Again students were assigned to condition as intact classes. Ten students from the Sourcer's Apprentice group and 19 students from the control group completed the entire project.

**Module.** The Sourcer’s Apprentice module used for this experiment concerned the Homestead steel strike. Seven authentic texts were provided to address the narrative question, “What happened in the summer of 1892 at the Homestead steelworks?,” and the controversy, “To what extent was Carnegie responsible for breaking the union at Homestead?” The excerpts were from a high school textbook, historian essays, participant accounts, and a congressional committee report.

---

Design. The design was the same as Experiment 2A.

Procedure. The procedure was the same as in Experiment 2A with minor differences. Due to a malfunction with several computers on the first day, students in the experimental group received the training and sourcing tutorial in groups of five. This was followed by a short lecture to provide necessary background information on Homestead so they could begin the instructional module at the beginning of the next class period. Students spent the following 40-min class period using the Sourcer’s Apprentice practice environment. Due to a shortage of computers, several students had to work in pairs. Finally, due to time constraints, their note cards were printed out and students were assigned the essay as homework.

Results and Discussion

The means for each test occasion (pretest vs. posttest) and condition (Sourcer’s Apprentice vs. control) are shown in Table 3. A mixed design ANOVA revealed no main effect for either condition, $F(1, 27) = 0.12, p = .73$, or test occasion, $F(1, 27) = 0.09, p = .77$. As in Experiment 2A, however, there was a significant interaction of condition and test occasion, $F(1, 27) = 6.42, p < .05$. The Sourcer’s Apprentice group improved by an average of 2.9 items from pretest to posttest, whereas the control group correctly answered an average of 3.7 fewer items on the posttest, again probably due to greater difficulty of the posttest topic (Panama Canal) relative to the pretest topic (Lexington–Concord). This yields an average relative difference of 6.6 (9%) more items correct by the Sourcer’s Apprentice group. Although the gain is smaller than that observed in Experiment 2A, it should be noted that due to computer difficulties the experimental group received training in groups of five and many had to work through the Sourcer’s Apprentice in pairs rather than individually. Indeed, considering that the students received only 40 min of actual time working on the module and that they were given the posttest after a weekend delay, these could be considered rather significant gains, especially given how notoriously difficult it is to find transfer effects. Optimally, students would have multiple opportunities to practice with the Sourcer’s Apprentice throughout the year on a variety of topics.

EXPERIMENT 3

In Experiment 3 we investigated whether the effectiveness of the Sourcer’s Apprentice was due to the nature of the environment or to the particular materials that students read. Many textbooks now cite primary and secondary sources and present excerpts as direct quotations or offsets in text boxes. This experiment
compared the Sourcer’s Apprentice against an integrated textbook treatment of the same materials. The Sourcer’s Apprentice group read a background text and engaged in a module in the same way as in Experiments 2A and 2B. The control group, however, received the same materials to study, but organized as in a textbook. The background text was the central text with the other seven documents inserted as quotations and cited in endnotes. If the two groups performed equally well, then we would conclude that the primary benefit of the Sourcer’s Apprentice was due to its exposing students to multiple sources of varying types, and such exposure could be much more efficiently provided through textbooks. If, however, the Sourcer’s Apprentice group performed better than the control group, then the effectiveness of the Sourcer’s Apprentice must be due to more than mere exposure to the document set, but to other factors such as how the documents were presented and the specific activities in which the students engaged.

Method

Participants. The same teacher who participated in Experiment 2A kindly provided us with access to two additional American history classes in early March of the following year. Nine participants from the experimental class and 14 from the control class completed the entire 4 days of the experiment.

Module. The module used was the Homestead module from Experiment 2B.

Design. Sourcer’s Apprentice and control groups were given the same materials, but in different ways. In the Sourcer’s Apprentice condition, the materials were document-centered. A background text provided an independent overview and then the documents were presented using the Sourcer’s Apprentice with a focus on individual documents on a bookshelf. In the control condition, the materials were textbook-centered (see Appendix). The experimenters modified the background text so that it included the seven documents in a single text. All significant source information was presented in the text. All content from the documents was presented either as an imbedded quote or in an endnote with a “[presented here]” statement in the text signaling the appropriate endnote. It is important to note that for the sake of coherence, an interpretation of the main point of the source was given when a reference was made to the source. Thus, although the Sourcer’s Apprentice group’s document-centered materials required integration of separate sources, the control group’s textbook-centered materials provided a single integrated version of the story for the students.

Procedure. On the first day all students completed the Panama Canal pretest (as described in Experiment 1). During the next class period, both groups
went to the library to learn about the Homestead steel strike, a topic in their normal curriculum. The control students (textbook-centered) worked in a library classroom without computers. They read the text and were able to take notes on a separate sheet of paper. After reading, students handed in the texts and kept their notes. Then they answered several source and content comprehension questions and were asked to write an essay of three to four paragraphs on the controversy.

The experimental class (Sourcer’s Apprentice) received the background text at the beginning of the first class period in the library computer room. Then they received a 5-min demonstration by the experimenter on how to use the program. Then students, working individually, went through a brief interactive tutorial on sourcing. When finished with the tutorial, the students read the seven documents and filled in the note cards for each document. Then, with only their note cards available, the students answered the same content and source questions as the control group. Finally, they were asked to write an essay and were given the same instructions as the control group. Students who did not complete the module in a single 35-min session and students who were absent the previous class period were given time to finish on the following day.

To motivate all students on the essay-writing task, they were told that if they included at least eight events or citations, they would be eligible for a chance to win a certificate for a music CD. One week later, all students took the posttest in their regular classroom and the teacher made no reference to the prior activities. The posttest was based on the Vietnam War module.

Results and Discussion

Because students have a tendency to ignore information presented in footnotes or endnotes, our analyses of both groups excluded references to material presented to the control group only in an endnote.

**Transfer test.** Pretest and posttest sourcing scores are shown in Table 3. A 2 (Condition) × 2 (Test Occasion) mixed ANOVA revealed a significant main effect of test occasion, $F(1, 21) = 20.43, p < .01$, reflecting an overall improvement in sourcing scores from pretest ($M = 9.2$) to posttest ($M = 15.4$). There was also a significant condition and test occasion interaction, $F(1, 21) = 14.76, p < .01$, due to larger posttest improvement for the Sourcer’s Apprentice group. Students using the Sourcer’s Apprentice showed an average gain of 14.1 points on the sourcing posttest, whereas the control group showed an average gain of only 1.1 points. This difference represents a 19% advantage for the Sourcer’s Apprentice group. Thus, presenting the same materials in an integrated, textbook format was not as effective as using the Sourcer’s Apprentice.
**Essay analysis.** Because both groups received the same documents and the same essay task, it is possible to compare the quality of the two groups’ essays. To analyze the essays, the information from all the texts was segmented into propositions, which were then used to provide an indication of the source of information. Both authors independently scored a random 40% of the essays. Since interrater reliability was acceptable (89%), the remaining essays were scored by only one author. All essays included at least the minimum of eight important events and took a stance on the controversy. Considering only material that was available in the text itself and omitting information in endnotes, we analyzed how much of the information mentioned by the students originated from the background text and textbook (narrative) sources compared to the historian accounts and primary documents (documents). The means for the mixed design ANOVA including condition (Sourcer’s Apprentice group vs. control group) and source of information (narrative vs. documents) are shown in Table 4. The analysis revealed no main effect of condition, $F(1, 21) = 0.216, p = .65$; a significant main effect of source of information, $F(1, 21) = 33.65, p < .01$; and a significant interaction of condition and source of information, $F(1, 21) = 8.36, p < .01$. Further post hoc tests on this interaction showed that the essays from the two groups mentioned an equal amount of narrative information ($M = 19.88$ vs. $M = 25.73$ for the Sourcer’s Apprentice group and control group, respectively), $t(21) = –1.15, p = .26$, but the Sourcer’s Apprentice group included significantly more document-based information than the control group ($M = 12.63$ vs. $M = 4.07$, respectively), $t(21) = 4.53, p < .001$. Presumably, presenting the material with a focus on individual documents, including making the source of the document salient, led students to use more source

<table>
<thead>
<tr>
<th>Source of information in student essays</th>
<th>Sourcer’s Apprentice-Centered</th>
<th>Textbook-Centered</th>
</tr>
</thead>
<tbody>
<tr>
<td>Narrative</td>
<td>19.88</td>
<td>25.73</td>
</tr>
<tr>
<td>Documents</td>
<td>12.63</td>
<td>4.07</td>
</tr>
<tr>
<td>Explicit references in student essays</td>
<td></td>
<td></td>
</tr>
<tr>
<td>References</td>
<td>3.38</td>
<td>0.47</td>
</tr>
<tr>
<td>Connectors in student essays</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Causal connectors</td>
<td>33.38</td>
<td>18.80</td>
</tr>
<tr>
<td>Temporal connectors</td>
<td>5.13</td>
<td>15.00</td>
</tr>
</tbody>
</table>

Table 4

Descriptive Statistics for the Information in the Student Essays for the Sourcer’s Apprentice and Control Groups in Experiment 3
information in essays than when the same information was presented in the context of a single author’s presentation.

This focus on individual documents also led to more explicit citations as shown in Table 4, \( t(21) = 4.59, p < .01 \). The essays from the Sourcer’s Apprentice group contained on average about three citations, whereas only one of every two control group essays contained a citation. In fact, every essay from the Sourcer’s Apprentice group included at least one reference. In contrast, only one third of the control group’s essays cited a reference.

Essay quality was also measured by a more traditional method: teacher grading. A high school history teacher (not the students’ teacher) was given the essays in a random order, blind to condition, and asked to grade the essays on a 5-point scale as a standard in-class assignment; that is, 4 = A, 0 = F. The average grade given to the essays was a B (\( M = 3.04, SD = 0.88 \)), ranging from eight A’s to one D. The Sourcer’s Apprentice group’s essays were given an average grade of 3.75 (\( SD = 0.46 \)), which was significantly better than the average grade of 2.67 (\( SD = 0.82 \)) given to the control group’s essays, \( t(21) = 3.44, p < .01 \). The substantive comments made by the teacher focused on number of details and specificity, the number and the quality of reasons to support stated positions on the controversy, and the citing of sources.

A final indicator of essay quality is the number and type of connectors used to integrate the content. Connectors were categorized as either temporal (e.g., “during the same time,” “afterward”) or causal (e.g., “because,” “due to,” “therefore”). The Sourcer’s Apprentice group’s essays contained more causal connectors than the textbook-centered group’s essays, \( t(21) = 2.58, p < .05 \), but there was no difference in the use of temporal connectors, \( t(21) = 0.08, p = .94 \). Although both groups read the same material, the group that used the Sourcer’s Apprentice wrote more causally integrated essays.

**Online data.** An analysis of the Sourcer’s Apprentice group’s online behavior provides some insight into how they performed the task and what they found difficult. A very consistent pattern emerged in the order in which students read and filled in source features even though the environment purposely provided the opportunity for considerable flexibility. There was a strong tendency for students to open and read the books in the left-to-right order in which they appeared on the bookshelf. In fact, 10 of the 12 students (83%) never deviated from the bookshelf order and 1 of the deviators read only one book out of order. A similar pattern appeared in the order in which students filled in the note card for each book. Recall that several source features were filled by a drag-and-drop procedure, whereas others were filled by the student typing a response in an input box. In general, students filled in the drag-and-drop features in the order in which they were presented on the screen. The sequence for filling in the features was generally in a top-to-bottom order: author features first, followed by document features (89%). The only major deviation was whether the content features were filled in immediately (56% of
occasions) or were skipped until later (33% of occasions). Thus, most of the time the students systematically read the books in the order provided and filled in the drag-and-drop features in their screen order.

Students filled in most of the drag-and-drop source features (94% of the 42 features; $M = 39.5$, $SD = 3.53$), but only occasionally filled in the insert source features (23% of the 21 features; $M = 9.5$, $SD = 6.6$). The task of identifying features was not trivial for the students. On average they made 7.58 errors per student ($SD = 4.52$). When they did have difficulty, they consulted the help screens 5.08 times ($SD = 4.34$) per student. Most of the help requests came before attempting to fill in a particular feature (92%) rather than in reaction to an immediate (2%) or previous error (7%). The help screens were useful, resulting in a correct answer immediately after closing a help screen 64% of the time. Most of the help requests occurred on the first two documents (64%), suggesting that students learned how to fill in each feature and stopped requesting help on subsequent documents.

A breakdown of student performance for each of the various features is presented in Table 5. The least frequently completed features were the three that required typing a response (Author Motives, $M = 4.33$; Main Point, $M = 4.25$; and Comments, $M = 0.92$), compared to an average completion rate for the drag-and-drop features of 6.59 out of a possible 7. Furthermore, students often skipped over the Author Motives to first complete the other drag-and-drop features. Of the drag-and-drop features, students had the most difficulty with how know (on which they made the most requests for help) and documents mentioned (on which they made the highest number of errors).

**Source feature recall data.** One week after using the Sourcer’s Apprentice, just after the posttest, the Sourcer’s Apprentice students were asked to recall as

<table>
<thead>
<tr>
<th>Feature</th>
<th>Average Number Filled</th>
<th>Proportion Recalled (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Who</td>
<td>7.00</td>
<td>55</td>
</tr>
<tr>
<td>Position</td>
<td>6.92</td>
<td>82</td>
</tr>
<tr>
<td>How Know</td>
<td>6.83</td>
<td>36</td>
</tr>
<tr>
<td>Author Motives</td>
<td>4.33</td>
<td>45</td>
</tr>
<tr>
<td>When</td>
<td>6.92</td>
<td>64</td>
</tr>
<tr>
<td>Type of Document</td>
<td>6.92</td>
<td>27</td>
</tr>
<tr>
<td>Documents Mentioned</td>
<td>4.92</td>
<td>18</td>
</tr>
<tr>
<td>Main Point</td>
<td>4.25</td>
<td>55</td>
</tr>
<tr>
<td>Comments</td>
<td>0.92</td>
<td>0</td>
</tr>
</tbody>
</table>
many of the source features as they could, with the reminder that some of the features were more pertinent to author, document, or content. The control group could not be given this test because they were not given the tutorial training. Overall, the Sourcer’s Apprentice students recalled 47.73% of the features ($SD = 22.41$), ranging from 1 student not recalling a single feature to 1 student recalling 75% of the possible features. As shown in Table 5, position (e.g., author credentials) was the best recalled (mentioned by 82% of participants), whereas document type (27%) and documents mentioned (18%) were the least recalled. Thus, posttest performance was likely limited by the students’ ability to recall some features. Presumably both recall accuracy and posttest scores would have been higher with multiple exposures to the Sourcer’s Apprentice over the term.

GENERAL DISCUSSION

In Experiment 1 we showed that students do not spontaneously attend to source information when studying and taking notes in a multiple-document, problem-solving situation, averaging only 15% to 23% correct on a sourcing test and citing as credible information obtained from a novel. This was true of both high school and college students, and it was true whether students were told to attend to source information or simply given comprehension instructions. Such blurring between legitimate historical data and fiction is consistent with Wineburg’s (1991) observation. More recently, Wineburg (2000) observed that both students and their parents make references to movies such as *Forrest Gump* and *Schindler’s List* to support claims in their discussion of history. Similarly, Seixas (1994) found that high school students considered the contemporary film *Dances with Wolves* as an authentic copy of the past, rather than a reflection of the late 20th century.

In Experiments 2A, 2B, and 3 we tested the effectiveness of the Sourcer’s Apprentice in teaching students to attend to source information when studying historical documents. In Experiments 2A and 2B we compared the Sourcer’s Apprentice to regular classroom activity. After the Sourcer’s Apprentice training, students mentioned more source features in their notes and answered more source questions correctly on a transfer test relative to students not given the Sourcer’s Apprentice training. We found this result for two different modules (Vietnam War and Homestead) and two very different populations of students (suburban and rural) after a relatively brief exposure to the Sourcer’s Apprentice.

In Experiment 3 we compared the Sourcer’s Apprentice to a control group that read a textbook-like version of the same training materials. In this textbook version, the students read an integrated presentation that provided interpretation for each source, as do many current textbooks. We found that training with the Sourcer’s Apprentice resulted in better sourcing performance than with the textbook version. As in Experiments 2A and 2B, students using the Sourcer’s
Apprentice did better on a transfer sourcing test. They also wrote better essays. Although there was no difference in the amount of narrative content in the essays, the Sourcer’s Apprentice group included significantly more document-based content and made more explicit references to sources in their essays. In terms of overall quality, the Sourcer’s Apprentice essays were given a higher grade by a history teacher. Thus, the Sourcer’s Apprentice with its focus on individual excerpts and sourcing seems to highlight secondary and primary source information. Interestingly, the results suggest that when documents are interpreted by an author in a single text, the documents are less likely to be used later in an essay. An implication of this finding is that if we want students to attend to primary and secondary excerpts, they should be presented as separate documents with limited interpretation by the focal source (i.e., the textbook or the lecturer).

The essay results from Experiment 3 are consistent with findings reported by Wiley and Voss (1999) showing that writing arguments from multiple sources presented in a Web-based environment resulted in greater transformation of the information in a subsequent essay than when multiple sources were not available. Such findings may also provide a possible reason for students’ tendency to ignore tables, figures, primary evidence, and text boxes in textbooks. If students think the important material has already been summarized, they have little incentive to spend time trying to understand it. A final interesting finding from Experiment 3 was that students trained on sourcing who read separate documents, that is, the Sourcer’s Apprentice group, tended to write better connected essays than students who read the same material in a textbook-like format.

Taken together, these studies show that many students are not adequately learning document-level skills even by the time they reach college, but that through a rather simple intervention, we can significantly improve their abilities. Although the improvements in sourcing ability were modest, so was the intervention. Students worked on a single problem for one or two short class periods. We expect that providing a more lengthy intervention would be even more beneficial for students. Furthermore, ours is an intervention that, according to evaluations by 8 high school and 6 college students, was enjoyable (median of 7.5 and 7.0, respectively, with 9 being wonderful), easy to use (median of 9.0 and 8.0, respectively, with 9 being very easy), and educational (median of 8.0 and 8.5, respectively, with 9 being very educational).

Furthermore, we believe that the Sourcer’s Apprentice could be extended to other domains and to other document-level skills. In our initial work with Sourcer’s Apprentice, we focused on the domain of history as a training ground for document-based learning and reasoning skills. From our work it is clear that document-level skills are generalizable to other domains that use documents as evidence and involve arguments to support theories, such as the humanities (e.g., philosophy, literature), the social sciences (e.g., economics, sociology), and the natural sciences (e.g., psychology, physics).
The Sourcer’s Apprentice could also be extended to other document-level literacy skills such as argument comprehension, information search, and integration. The necessity of learning such skills is more urgent now than ever before as technology expands access to information (Britt & Gabrys, 2000). The Internet has reduced the cost of publishing to virtually nothing, flooding the marketplace of ideas. At the same time, those whom society has traditionally relied upon as gatekeepers of credibility, such as editors and publishers, are facing increased competition and decreased time to print, leading to less scrutiny of sources. All of this places students in the position of needing to filter and evaluate information sources. Applications such as the Sourcer’s Apprentice are one way to give students the skills they need.

ACKNOWLEDGMENTS

This study was financially supported by the James S. McDonnell Foundation. We would like to thank Samuel Wineburg, Deborah McCutchen, and Jean-Francois Rouet for their helpful comments on an earlier version of this article. We also wish to thank Nicole Hunka and Julie Van Dyke for help in running one of the experiments. We are very grateful for Mara Georgi’s help in module development. Finally, the Sourcer’s Apprentice development team included Charles Perfetti, Mara Georgi, and Gareth Gabrys.

REFERENCES

Birnbaum, L. (1986). Red dawn at Lexington: If they mean to have a war, let it begin here! Boston: Houghton Mifflin.


In 1892 the Carnegie Steel Company cut wages at its Homestead, Pennsylvania, plant. In July the AAISW—the strongest AFL union—called a strike. Instead of bargaining with the union, plant manager Henry Clay Frick decided to kill it. He locked the workers out of the plant. Then he hired 300 armed guards from the Pinkerton Detective Agency to enforce the lockout and to protect the new, nonunion workers he had hired to replace the strikers. When the Pinkerton guards arrived, the unionists were waiting for them. Violence broke out, and seven guards and nine steelworkers were killed. The guards left, but the governor of Pennsylvania sent 8000 militia to maintain order. The unionists watched as new workers arrived to take their jobs. The strike dragged on until November, but by then the union was dead and thousands of workers had lost their jobs. (King, McRae, and Zola, 1993)5

These are the historical events that occurred in Homestead just before the turn of the century as described by a popular history textbook. But how did things get so out of hand that violence and the death of a labor union was the result? The following pages will describe the background of this conflict so you can better understand these important events.

Business in the Late 1800s

In the years after the Civil War (which ended in 1865) industrialization spread rapidly in the northern part of the United States. In most cities, huge factories were built to make everything from steel to sewing machines. Business owners began to be more concerned with earning a large profit and keeping their expenses low. This changed the relationship between a boss and the workers. With factories using strong managers, workers were subjected to new rules.

Factory Conditions

Factory owners were not required to make their factories safe so many workers were injured or killed by the equipment. Also, most industries expected their workers to...
work 10 hours a day for 6 days of the week. For this 60 hr work week, the average yearly income was between $400 and $500. At that time, the minimum amount of money needed to live at a reasonable level of comfort was close to $600 per year.

Factory workers had no control over any aspect of their working lives. They were given only a brief lunch break. Managers did not even give workers a chance to stop and use the bathroom. Workers who complained about the bad conditions were generally fired or blacklisted. Blacklisted means that the manager told other factories not to hire the worker because he was a “trouble-maker.” Therefore, workers usually had to tolerate the conditions because they risked losing the job they had and any possibility of being hired elsewhere.

Formation of Unions

Low pay, long hours, dangerous machines, no government regulations, and a lack of control all led to worker discontent. Eventually, the workers joined together in an attempt to fight the horrible factory conditions. They joined together in unions. One of the first unions was called the Knights of Labor. For a short period, these unions won higher pay and shorter work hours for their members. Unfortunately, there were no laws protecting workers and soon the strikes were crushed by the owners. In two famous cases, the Homestead steel strike in Pittsburgh and the Pullman railroad strike in Chicago, the unions held large strikes that might have been powerful enough to win higher pay for workers. However, in both cases police and national guard units were called in to protect replacement workers. Also, in both cases the strike leaders were arrested and jailed. The strikes failed, workers were forced to accept lower wages, and owners were again allowed to do whatever they wanted. These two major failures basically destroyed the American Labor movement.

The Homestead Steel Strike of 1892

One example of the clash between owners and workers took place in Homestead, Pennsylvania in 1892. On one side was the union, the Amalgamated Association of Iron and Steel Workers (AAISW). On the other side was the management of the steel mill. The owner was Andrew Carnegie and his general manager Henry Clay Frick.

The steel mill. The Massive steel mill employed 3800 men. The majority of the workers were nonskilled who were paid by the skilled workers who actually made the steel bars and plates. These skilled workers were paid under a complex system called the sliding scale. As the price of steel rose, the workers earned more. However, if prices sank below a certain level, wages would not fall any
further. This level was known as the minimum level. If wages fell below this minimum, the workers would not be able to earn a living. In 1892 the minimum was $25.00 per ton.

**Events leading up to the strike.** Negotiations for a new contract began in January of 1892. Their previous contract was to end July 1st and Frick decided that the AAISW had to accept the new contract by June 24. The union presented its side: continue working for the same wages, and continue having the union represent the laborers. Management wanted to make changes to the contract: lower wages and a lower minimum for the sliding scale ($22 per ton).

In May, Carnegie departed for vacation in Scotland and left Frick in charge. As talks continued, management communicated with the Pinkerton Detective Agency to protect the steelworks in case of a strike. The Pinkertons were hired during strikes because they supported managements’ attempts to break unions by protecting nonunion workers who were hired to replace the striking workers.

The final negotiating session was held on June 23, 1892. Both management and the union held firmly to their demands and so very little bargaining was done. The negotiations failed. Then the company shut two departments and locked out approximately 800 men. This action prompted the workers to hold another mass meeting at which they decided to refuse to work with replacement workers.

**The strike.** In reaction to the workers’ refusal to work, the company closed the entire mill and declared that it would no longer recognize the union. On July 5th, the sheriff failed in his attempt to open the mill to replacement workers. So management tried to secretly get the Pinkertons into the mill by sailing them up the Monongahela River and landing on the steel mill’s property. The workers spotted them and rushed to prevent the Pinkertons from landing. For an entire day, the Pinkertons and workers fought. Finally, the Pinkertons surrendered to the Homestead sheriff.

Management then moved the National Guard into Homestead and successfully secured the steelworks and reopened the mill. Management set a deadline of July 21 for the workers to reapply for their jobs without union representation. But they continued to strike and the management happily hired new workers.

**Responsibility for the strike.** There is disagreement over who in management was most responsible for the strike, Carnegie or Frick. Samuel Yellen is an history professor at Indiana University who believes that Frick provoked the strike. In his 1936 book entitled *American Labor Struggles*, he wrote:

> When Henry Clay Frick was given the managing authority of the company, the workers realized that they would have to fight for the preservation of their union
against the avowed and ruthless anti-union policy of the man who had already crushed several strikes by means of the Coal and Iron Police, the Pinkerton Detective Agency, and the state militia.

Before the appointment of Frick the men had believed somewhat in the friendship of Andrew Carnegie, poor immigrant boy from Dunfermline, donator of libraries and hospitals and music-halls, patron of the workingman and democracy and peace. He had written frequently concerning the relations of capital and labor; he had advocated trade-unionism and the peaceful arbitration of differences, and had deplored absentee capitalism and the violence of dispute.

Negotiations for a new agreement between the company and its employees opened in February, 1892, when a committee from the Amalgamated presented a scale to J.A. Potter, superintendent of the Homestead mill. Potter, in turn, handed the committee a scale from the company, providing for a reduction in wages and calling for a change in the date for the termination of the contract. The committee asked why a reduction was demanded, but no explanation was forthcoming.

The men haggled with the company for the next three months, but the conferences brought no accord. Finally on May 30 the company issued an ultimatum that the men would have to accept its scale before June 24. After that date the men would be dealt with only as individuals. […] Meanwhile alarming preparations were being made at the Homestead works. A solid board fence was erected around the mill property. This fence was topped with barbed wire and was perforated at intervals, as if for rifles, although Frick testified before the Congressional investigating committee later that the holes were meant for observation; and in the mill-yard stood platforms equipped with searchlights. The workmen dubbed the mill Fort Frick.

At the request of the workers Frick met a committee from the Amalgamated in the company office at Pittsburgh for a final conference on June 23, the day before the expiration of the ultimatum. […] Throughout the conferences the men had been offended by the cold uncompromising attitude of the company and particularly by the fortifications thrown around the works even before the negotiations were at an end. Now they were angered by the obvious aggressiveness of the company, and understood fully that behind all the differences had been the single question of the preservation of their union.

The company attitude rankled all the more because the men prided themselves on their Americanism and on the conservative policy and reasonable spirit of the Amalgamated. In this feeling the 1893 U.S. House of Representatives Congressional investigating committee later concurred with the men [presented here]. 6 In their

---

6 As quoted from the U.S. House of Representatives Congressional Committee 1893 report:

We conclude from all the surroundings that Frick, who is not the only manufacturer thus affected, is opposed to the Amalgamated Association and its methods, and hence had no anxiety to contract with his laborers through that organization, and that this is the true reason why he appeared to them as autocratic and uncompromising in his demands. If, as he claimed, the business of his company, on account of fall in the market price of the products of the works, required a reduction of the wages of the employees, he should have appealed to their reason and shown them the true state of the company’s affairs. We are persuaded that if he had done so an agreement would have been reached between him and the workmen, and all the trouble which followed would thus have been avoided. (pp. 449)
report they found that Frick had patently tried to force the workmen into opposition, and he had succeeded.\textsuperscript{7}

Yellen was not alone in arguing that Frick was responsible the events at Homestead. Even owner Carnegie suggests in his 1920 autobiography that it was Frick who was responsible.

I was traveling in the Highlands of Scotland when the trouble arose, and did not hear of it until two days after. Nothing I have ever had to meet in all my life, before or since, wounded me so deeply. No pangs remain of any wound received in my business career save that of Homestead. It was so unnecessary. The men were outrageously wrong. The strikers, with the new machinery, would have made from four to nine dollars a day under the new scale—thirty per cent more than they were making with the old machinery. While in Scotland I received the following cable from the officers of the union of our workmen:

“Kind master, tell us what you wish us to do and we shall do it for you.”

This was most touching, but, alas, too late. The mischief was done, the works were in the hands of the Governor; it was too late.\textsuperscript{8}

Other scholars find Carnegie was most responsible for provoking the strike. Paul Krause, history professor at the University of British Columbia, holds this belief. In his 1992 book entitled \textit{Battle for Homestead, 1880–1892: Culture, Politics, and Steel} Krause wrote:

One person remains curiously invisible on this important day in American history. Just where was the chief officer of Carnegie Steel while his Pinkertons and his steel-workers were locked in battle on the banks of the Monongahela?

\textsuperscript{7}Samuel Yellen is a Professor of History at Indiana University. He received his MA from Oberlin College in 1932. In preparing this classic of labor history, Yellen studied extensively ten historical confrontations between working men and women and the owners of America’s mines, mills, and railroads. He attempted to analyze the tactics and weapons of industrial warfare used by both workers and management. These include strikes, sabotage, lockouts, and the hiring of armed guards. He was interested in how these tactics changed from 1877 to 1934.

\textsuperscript{8}Andrew Carnegie was born in Scotland and moved to the United States in 1884. Although he arrived with no trade and very little money, he quickly invested in the growing iron industry. Within 20 short years he was one of the wealthiest men in the world. He then spent the last 18 years of his life giving away approximately $350 million. This excerpt is about the events during the Homestead Strike where he was the owner of the Homestead Steelworks. As the owner, he was responsible for the management of the steelworks facilities and workers. In this excerpt he tries to explain and justify his actions during the critical events.
On 5 July, as the Pinkerton detectives were making their way to Homestead, Andrew Carnegie, far from the din of his steelworks, was being happily cheered in Aberdeen, Scotland, where he had come to open formally the library he was bestowing on that municipality. In appreciation of this latest installment in Carnegie’s burgeoning empire of philanthropy, he received the freedom of the city and was lavishly hailed by local officials.

Carnegie spent the night in Aberdeen at the Haddo House hotel, where he received a telegram from Pittsburgh notifying him of the battle at Homestead. Undeterred by the news, he proceeded with his vacation plans and left on 6 July for Rannoch Lodge, a retreat in the central highlands, so isolated that it could be reached only by private carriage. Loch Rannoch and the nearby streams provided Carnegie with an idyllic setting to pursue a favorite pastime, fishing. While vacationing at the lodge, he received a number of urgent telegrams from America, [. . .].

In truth, the Homestead Steel Works was hardly a recent acquisition. Carnegie had owned it for nearly a decade; moreover, he had pursued a vigorous policy of antiunionism there, as elsewhere throughout his empire of steel. To be sure, in April 1892, well before leaving the United States for his vacation in Scotland, Carnegie had sent Frick the draft of a notice [present here]9 meant for the Homestead workers which stated that the company would no longer recognize any union upon expiration of the contract on 30 June. In a subsequent letter to Frick, dated 4 May, Carnegie reaffirmed this plan and also, on behalf of the senior partners of the firm, granted his chief of operations full authority to proceed as he saw fit.

Given his physical distance from the events and a convincing attitude of shock and dismay, Carnegie could claim somewhat credibly that he had no direct role in the violence that shook Homestead. It was, after all, Frick who, in Carnegie’s absence, had called in the Pinkertons. (In this he was following a precedent established by Carnegie in a dispute at Braddock in 1888.) However, in creating and fortifying the system that had, over the years, produced the conditions for this violence, Carnegie’s role cannot be denied. What provoked the apparent “barbaric” and “thankless” workers of Homestead was not, as an account limited to that day might indicate, the sudden intrusion to [sic] Pinkerton agents into their dispute but the slow and steady erosion of their rights and their power, over which Carnegie and

---

9Carnegie’s Notice to employees at Homestead Works was written on April 4, 1892. He planned to be out of the country during contract negotiations and he wrote this letter just before he left the country. He intended on having it posted, if necessary, in late June as an explanation of why he wanted to end the union.

These Works having been consolidated with the Edgar Thomson and Duquesne, and other mills, there has been forced upon this Firm the question Whether its Works are to be run ‘Union’ or ‘Non-Union.’ As the vast majority of our employees are Non-Union, the Firm has decided that these works, therefore, will be necessarily Non-Union after the expiration of the present agreement. [. . .]

This action is not taken in any spirit of hostility to labor organizations, but every man will see that the firm cannot run Union and Non-Union. It must be either one or the other.

It was reprinted in Demarest and Weingartner (1992, p. 204).
his associates in steel and politics had presided for years, invisibly but no less violently.\(^{10}\)

Krause’s position is further backed up [sic] by the recollection of Carnegie’s private secretary, James Bridge. Bridge wrote in his 1903 memoirs *History of Carnegie Steel Company* that Carnegie was responsible for the bloody events in Homestead:

> There was one spot, however, where these items of news did not readily penetrate; and that was Rannoch Lodge. Here, thirty-five miles from the nearest railway and telegraph station, Andrew Carnegie, in accordance with plans previously made, denied himself to reporters and refused to answer telegrams or letters relating in any to Homestead. Having delegated his authority to Mr. Frick, he knew that the measures they had jointly planned would be carried out to the letter, despite the efforts of anarchists or the protests of politicians. And so he went fishing; and the London papers sought in vain to get an expression of opinion from him on the Homestead battle.\(^{11}\)

**The strike collapses.** The AAISW did not give up the strike until November 18 when a majority of the workers returned to work and accepted management’s contract. This included a large pay cut. For example, the boilerman who earned $2.25 per day before the strike was paid $1.89 daily after he returned to work. Carnegie Steel had won so completely that it would be over 40 years before steelworkers in Pennsylvania again attempted to unionize.

\(^{10}\)Paul Krause is an Professor of History at the University of British Columbia in Vancouver Canada. He received his PhD from Duke University in 1987. Krause is a historian who specializes in 19th century U.S. social and labor history. He has written several academic publications on the history of labor in America and the Homestead Lockout in particular. He grew up in Pittsburgh in the 1950s. Although his family was not steelworkers, his childhood was greatly influenced by the great steel mills of the Monongahela Valley. He believed that the reality and romance of steel are important in understanding America’s development as a Nation in the 19th century.

\(^{11}\)James Bridge’s memoir’s were originally published in 1903 under the title, *History of Carnegie Steel Company*, by Aldine Book Co, New York. It was later reprinted in *The inside history of the Carnegie Steel Company*. James Bridge was Andrew Carnegie’s private secretary and was in Pittsburgh during the Homestead Steel strike. He strongly favored business interests over labor concerns. Around 1901, Bridge left Carnegie over a major disagreement about another employee. Bridge then spent the rest of his life criticizing Carnegie’s accomplishments.