Students’ Sourcing While Reading and Writing From Multiple Web Documents

Helge I. Strømsø
Professor, Department of Education, University of Oslo, Norway
h.i.strømsø@ped.uio.no

Ivar Bråten
Professor, Department of Education, University of Oslo, Norway
ivar.braten@iped.uio.no

ABSTRACT
This study focuses on how and to what degree undergraduates attend to source information while reading a set of partly contradicting web documents on a social-scientific issue, and on how students use those sources in writing an essay based on the documents. Think-aloud data collected during the reading session showed that students to some degree pay attention to both source information about the documents and to information about sources embedded in the documents’ content. Sourcing of an evaluative nature occurred most frequently while students read documents holding strong opinions on the issue. In the essays, students most frequently cited the source assumed to be most trustworthy. Students’ explicit online sourcing while reading was significantly related to their explicit references to the web documents in the essays. The results from the study are in accordance with prior studies in indicating that undergraduates need more training in sourcing skills.

Keywords
Sourcing, multiple documents, digital literacy, critical reading

INTRODUCTION
The Internet has forever changed the landscape of our daily life. We use the Internet to deal with a number of different tasks at work, at home, and at school. On a daily basis 80% of the Norwegian population access the Internet, and of those 73% are searching for different kinds of information (Statistics Norway, 2013). The Internet has become a main source of information to the majority of Norwegian citizens. The amount of available web-based information has increased immensely during the last decade, while standards for publishing have blurred. The recent digital information boom has partly shifted the responsibility for judging the quality and trustworthiness of information sources from editors and publishing houses to users of the Internet. It seems
reasonable to assume that skills in evaluating source features of documents retrieved from the Internet are important in order to avoid information overload and use of inaccurate information. In the present study, we explored to what extent undergraduates demonstrate such sourcing skills in reading and writing from multiple web documents on a controversial topic.

Reader attention to, as well as evaluation and use of, source information, are of importance in many different tasks. Such sourcing skills are deemed essential features of academic craftsmanship, and they may also be important when solving daily tasks outside the academic sphere. There are reasons to believe, moreover, that sourcing skills may facilitate student learning.

The term “sourcing” refers to people’s explicit attention to source information, their use of source information in predicting and interpreting a document’s content or judging its trustworthiness, and their citing of sources in written essays (Wineburg, 1991; Britt & Gabrys, 2001; Rouet, 2006). Sourcing implies actions taken by readers to identify the source of a message and use source information in interpreting and evaluating a text. For writers, sourcing implies justifying a statement by source information and following academic conventions in giving credit to the source of a particular message. Although attention to and use of source information characterize sourcing in both reading and writing, sourcing thus implies different activities in reading and writing. For the sake of simplicity, we will use only “sourcing” when referring to reading activities and “citation of sources” when referring to writing activities. Student attention to and use of source information are important in both reading and writing for several reasons.

**Theoretical framework**

Reader attention to, memory for, and use of source information have been studied within different theoretical frameworks, such as social psychology (Pornpitakpan, 2004) and memory research (Johnson & Mitchell, 2002), whereas the Documents model, proposed by Perfetti, Rouet, and Britt (1999), and further developed by Britt and Rouet (2012), specifies the role of source information in student reading and comprehension of multiple documents. The Documents model extends the description of mental text representations within Kintsch’s (1998) Construction – Integration model (CI) of single-document comprehension. With the Documents model focusing on multiple-documents reading, two layers are added to the CI model. The Situations model represents an integrated mental representation of the content in the different documents, while the Intertext model represents source information and relationships between the different documents. Given the focus of the present study, the Intertext model is of particular interest.

In regard to the Intertext model, Britt and Rouet (2012) emphasize that documents are not only linguistic expressions, but also constitute social entities, with source information indicating a document’s purpose and status in a
social context. Expert readers will often take such information into consideration in deciding how much effort they should invest in processing a document, how to process the document, and how to interpret the document’s content (Bazerman, 1985; Wineburg, 1991), whereas non-experts seem to be much less concerned about the role of source information (Britt & Aglinskas, 2002; Wineburg, 1991). Source information may include details about the author (e.g., name, credentials, personal ties) and the document (e.g., document type, publisher, date and place, cultural context), and it is assumed that such information becomes tagged to core content units in skilled readers’ representations of the documents. By tagging source information to important content units the reader will be able to keep track of where different content units come from and how they are interrelated.

Although more empirical studies are needed to support the Intertext model described by Britt and colleagues (Britt & Rouet, 2012; Perfetti et al., 1999), the model has contributed to our understanding of the role of source information in multiple-documents reading, and it represents an appropriate point of departure for further research on student sourcing skills. It is possible, however, that the model needs to be extended to include additional layers of source information. For example, expository documents often refer to other sources in the running text, which may in some cases, cause the reader to pay more attention to cited sources than to information about the document source (Strømsø, Bråten, Britt, & Ferguson, 2013).

One should also note that although the Intertext model primarily represents sourcing as an aspect of reading, student citations in written essays can also be described as sourcing skills (Britt, Wiemer-Hastings, Larson, & Perfetti, 2004), yet are not explicitly accounted for by the model. One could argue, however, that an essay is a kind of external reflection of the reader’s mental representation of the documents read, and that citations in an essay therefore indicate, at least partly, the reader’s Intertext model.

**Prior research on the processing and representation of sources**

Spontaneous sourcing refers to the processes of paying attention to, reflecting on, and using source information in interpreting the content of multiple documents while reading them. In Wineburg’s (1991) think-aloud study, focusing on differences between expert historians and high-school students reading multiple documents, results showed that expert historians intentionally sought out source information for each document while high-school students often ignored such information. In contrast to the high-school students, experts not only took note of source information, but also used that information actively in interpreting, evaluating, and predicting the content of documents. Results from later studies on spontaneous sourcing in multiple document reading, also indicate that both high-school students and undergraduates show low interest in source information and do not consider such information helpful for interpretative or evaluative purposes (Britt and...

Other studies have focused on how source information is represented in reports and essays written after the reading of multiple documents, that is, in products of reading processes. Results from these studies indicate that both the nature of the documents (primary vs. secondary) and the readers’ academic experience (undergraduates vs. graduates) might affect student representation of source information in essays. Undergraduates reading document sets including primary documents had a higher frequency of citations in essays than students not reading primary sources (Rouet, Britt, Mason, & Perfetti, 1996), and graduate students included more citations than undergraduates (Rouet, Favart, Britt, & Perfetti, 1997). In general, undergraduates reading multiple documents do not frequently cite those documents while writing summaries or essays (Britt et al., 2004; Le Bigot & Rouet, 2007; Rouet et al., 1996). Although citation of sources in essays cannot be interpreted as a complete external representation of a student Intertext model, the results are consistent with the view that students in higher education generally do not develop a very rich Intertext model while reading multiple documents. Also, a study by Le Bigot and Rouet (2007) suggests that students may rather refer to sources embedded in document content than to source information about the documents themselves.

The present study

Reasons for focusing on student sourcing and citation skills may be normative or related to student processing and understanding of multiple web documents. Normative reasons concern student adoption of citation conventions in the academic world. However, students also need to develop sourcing skills to evaluate whom or what to trust when different sources present conflicting information. Finally, as suggested by the Documents model, student attention to and memory for source information may facilitate multiple-document comprehension (Perfetti et al., 1999). Results from prior studies indicate that students at high school and undergraduate levels do not pay sufficient attention to source information while reading multiple documents, and that they only cite sources while writing essays based on the reading of multiple documents to some extent (Britt and Aglinskas, 2002; Le Bigot & Rouet, 2007; Rouet et al., 1996).

In the present study, we set out to investigate undergraduate student sourcing while reading multiple documents in unprecedented detail, through the use of think-aloud methodology. In addition, we wanted to study possible relationships between student sourcing while reading and their use of citations in subsequent essays. An explanatory study was designed to examine students’ spontaneous sourcing while reading multiple web documents on a controversial topic – cell phones and potential health risks.
Based on prior research in the area, the study addressed four research questions. The first question concerns the extent of spontaneous sourcing while reading multiple web documents on a controversial topic, and whether such sourcing only implies attention to source information or also evaluation of such information. Results from prior studies indicate that undergraduates display sourcing to a limited degree while reading multiple web documents (Stadtler & Bromme, 2007).

Second, does student attention to different source features (e.g., author, date, publisher) vary for different documents while reading? Prior studies have indicated that students evaluate the trustworthiness of multiple documents according to different source features after having read the documents (Bråten, Strømsø, & Britt, 2009; Rouet et al., 1996). We know of no other studies reporting on spontaneous sourcing in terms of student attention to source features, but assume that such attention could vary considerably across documents.

Third, to what extent do students cite sources when writing an essay based on their reading of multiple web documents? Based on prior studies (Britt et al., 2004; Le Bigot & Rouet, 2007) we expected the majority of students to include citations, but only a few explicit citations, with a substantial portion of the citations referring to sources cited in the documents rather than to the sources of the documents themselves.

Finally, we wanted to explore whether there is a relationship between students’ spontaneous sourcing while reading multiple web documents and their use of citations in essays. We expected that students would need to pay attention to source information while reading in order to cite documents when later writing an essay based on the documents.

METHOD

Participants

Participants were 51 students at a large state university in southeast Norway, who followed a mandatory introductory course in educational science at the bachelor level. The sample included 43 females and 8 males, with mean age of 22.1 (SD = 2.6).

Materials

Topic knowledge measure

Participant topic knowledge was assessed by asking them to respond in writing to three open-ended questions concerning what they knew about: (1) how cell phones send and receive signals; (2) possible health risks when using cell phones; and (3) how researchers investigate whether cell phone use might involve health risks. Each answer was scored from 0 to 3, based on specific criteria concerning what information it should contain. A random selec-
tion of 24 students’ responses to the three questions was independently scored by the two authors, resulting in a 90.3% agreement on the scoring.

Documents

Participants read six web documents presenting different perspectives on cell phones and potential health risks. The documents represented different kinds of authentic source materials. The following documents were included:

Science textbook excerpt. One 554-word document was excerpted from a traditional textbook in natural studies for upper secondary education. This document explained the functioning of cell phones and electromagnetic radiation in relatively neutral, academic terms, as well as how researchers can investigate potential health risks related to cell phone use.

Public information text. This was a 444-word public information text from the National Radiation Protection Agency’s website (NRPA), stating that the claim that cell phone use may cause serious illness is thus far not substantiated by research, but that some precautions should still be taken given a lack of longitudinal studies and studies involving children.

Popular science article. This text was a 473-word article from the website of a popular science magazine called Illustrated Science, where a research reporter described a recent, yet unpublished review article by an Australian professor and brain surgeon, Dr. Khurana, who argued that there is solid scientific evidence for the link between cell phone use and certain brain tumors.

Debate article. This document was a 483-word debate article from the website of an independent trade journal called Today’s Medicine written by a chief engineer representing the Cell Phone Operators’ Association, who took issue with Dr. Khurana’s message and argued that, given the current evidence, the cancer risk associated with cell phone use is exaggerated.

Subjective magazine article. This article was a 464-word text written by a reporter in a U.S. power critical web-magazine called the Idaho Observer, generally emphasizing advocacy journalism and populist political issues. The article stated that cell phone use doubtlessly causes cancer, citing scientific journals and researchers in support of this view.

Newspaper article. The last document was a 395-word online newspaper article written by a journalist in a Norwegian liberal daily, presenting an interview with a musician diagnosed with a brain tumor who suspected that his heavy cell phone use might have been the cause of his disease.

Apart from the more neutral science textbook excerpt, the other documents contained partly conflicting information. The popular science article, the debate article, and the subjective magazine article represented the strongest
opinions on the issue, with the debate article opposing and explicitly referring to the viewpoints presented in the popular science article, while the subjective magazine article contained no explicit references to specific information in the other documents. The public information text and the newspaper article represented more moderate stances, also reflecting that the issue remains unsettled.

The debate article and the subjective magazine article each cited 10 different sources, while the popular science article cited eight sources. The newspaper article cited two sources and the textbook excerpt and the public information text each cited one source.

In a prior study (Bråten, Braasch, Strømsø, & Ferguson, in press), participants were asked to rank the same six documents according to trustworthiness. The textbook excerpt was ranked as significantly more trustworthy than the five other documents and the public information text was ranked the second most credible document, significantly more so than the rest. There were no significant differences between the trustworthiness rankings of the four other documents.

Essay task
After reading all six web documents, participants were asked to write brief essays on the issue discussed in the documents. Specifically, they were given the following written instruction: “You are now going to prepare a brief report where you judge the health risk of cell phone use. Base your report on information from the texts that you have just read and try to express yourself clearly and elaborately — preferably in your own words. Justify your conclusions by referring to the sources you have been working with.”

Procedure
Participants were recruited to the study at a lecture intended for all students enrolled in the educational science course. The 51 students who chose to participate completed the reading task individually in a university lab, directed by a trained research assistant. Following Ericsson and Simon (1993), participants were instructed to verbalize all their thoughts and actions during reading. If they stopped talking for more than 30 seconds, the research assistant prompted them by reminding them to think aloud.

Before they started to read, participants were told that a fictitious friend experienced discomfort when using her cell phone, and that she had asked for advice. They were also told to imagine that they had already searched the WWW for information about the topic, and that the search resulted in six documents that they would like to take a closer look at. They could read the documents in any order they preferred and could go back and forth between the documents. Participants then entered a browser showing a Google-like
search results page, with links to the documents listed beneath the heading “search results”. All verbalizations and moves were recorded and videotaped by the usability software Morae (© TechSmith).

After reading all six web documents, participants were instructed to write brief essays based on the information they had encountered. They were not given access to the six documents while working on the essays because we wanted them to write their essays on the basis of the mental representations constructed during reading rather than on the basis of searching for, locating, and copying information at the time of performing the essay task. They were given maximum 20 minutes to complete the essay task.

Data analysis

Verbal protocols

Recordings of think-alouds were transcribed, resulting in 51 protocols containing 74039 words ($M = 1451.75$, $SD = 1059.89$, max = 5546, min = 162). From a total of 3658 utterances we identified 1103 units of sourcing in the transcripts.

Explicit sourcing included statements where participants clearly expressed that they had noticed or used source information by verbalizing characteristics of the source, such as name of publication, document type, author, or date in a precise way. Statements that indicated sourcing without characteristics of the source being verbalized in a precise way, that is, the way it was named in the documents, but still unambiguously referred to source information in the document set, were labelled implicit sourcing.

Next, we distinguished source information about one of the six documents in the document set, labelled present source, and source information included in the document content, that is, embedded source. When participants referred to source information about one of the six documents, but not the one they were currently reading, such sourcing was labelled other source in document set.

Finally, sourcing was coded according to the kind of sourcing activity the segment indicated. Partly based on prior research (Bazerman, 1985; Wineburg, 1991), and partly on preliminary readings of the protocols, we used two categories. When participants only mentioned source information without any further elaboration this was coded as attention. When participant sourcing involved any kind of evaluation of the trustworthiness of the source and its content, this was coded as evaluation (see Appendix 1 for coding categories).

The two authors coded a random selection of 25% of the protocols independently and reached an agreement of 84.6% in coding the units into the categories described above.
Essays

The mean number of words in the essays was 235 ($SD = 57.7$, max $= 379$, min $= 124$). Sentences containing source information referring to the six documents were defined as units of analysis. As stated above, references to source information were coded as either explicit or implicit. Explicit citations required the inclusion of accurate source information, such as name, publisher, or author. Implicit citations indicated beyond doubt that students were referring to sources in the document set, although such citations lacked specific source information (e.g., “The document that describes how cell phones work...”). Next, units including citations to one of the six documents were coded as present source, while units including citations referring to sources in the documents’ content were coded embedded source. Units including citations referring to several documents in the document set were coded general. Such citations could include general terms like “the documents”, but also cases where the same embedded source was included in more than one document and it was impossible to distinguish which document students were citing in the essays. The latter type of citation would be coded as general and explicit, while all other general citations would be coded as implicit.

Again, the two authors independently coded a random selection of 25% of the essays into the described category system, with this resulting in an agreement of 88.7%.

RESULTS

Topic knowledge

In sum 83% of the participants obtained a score of 3 or below on the topic knowledge measure and no one scored above 5 ($M = 2.51$, $SD = 1.27$). The results thus indicated that the participants could be regarded as laypersons with respect to the topic, and topic knowledge scores were not included in further data analyses.

Students’ spontaneous sourcing while reading

The mean number of source-related utterances produced by the participants while reading was 21.65 ($SD = 21.15$). With a minimum of 1 and a maximum of 112, there was a large variation in student sourcing activity. The mean number of explicit sourcing ($M = 12.63$, $SD = 13.66$) was slightly higher than implicit sourcing ($M = 9.02$, $SD = 8.41$), with a Wilcoxon Signed Ranks test indicating that the difference was statistically significant ($z = 2.43$, $p < .05$, $r = .24$).
The mean number of sourcing utterances to *present source* across the documents was 8.30 (SD = 9.59). Figure 1 shows that students produced the highest frequency of *present source* utterances for Idaho Observer ($M = 2.24$, $SD = 2.80$) and Today’s Medicine ($M = 2.03$, $SD = 2.79$), with references to Today’s Medicine being statistically significantly higher ($z = 2.23$, $p < .05$, $r = .22$) than references to the third highest, NRPA ($M = 1.41$, $SD = 1.76$). One should note that Idaho Observer and Today’s Medicine represent strong and opposing stances on the issue of cell phone use and health risks. However, as the Illustrated Science document also presented strong opinions on the issue, and explicitly contradicted Today’s Medicine, one might also have expected a high frequency of sourcing for that document. This is indeed the case, but not for utterances referring to *present source*. Instead, the frequency of references to *embedded sources* is higher for Illustrated Science than for any other document, although students also referred frequently to *embedded sources* in Idaho Observer and Today’s Medicine. The relatively high numbers of references to *embedded sources* in those documents were probably due to the number of citations included in the documents.
The frequency of sourcing in the form of evaluation indicates that the three documents including the strongest opinions and displaying the most explicit contradictions also elicit more sourcing of an evaluative nature than the other three documents (see Figure 2). A closer look at the data, however, reveals differences in the frequency of evaluation for present source, with the mean amount of evaluative sourcing being 1.00 ($SD = 1.85$) for Idaho Observer, .36 ($SD = .69$) for Today’s Medicine, and .29 ($SD = .64$) for Illustrated Science. It thus seems that the majority of evaluative sourcing concerning Idaho Observer is directed at the present source, while the majority of evaluative sourcing concerning Today’s Medicine and Illustrated Science is directed at embedded sources. This may be explained by the advocacy style of the Idaho Observer and the fact that both Today’s Medicine and Illustrated Science frequently refer to another source (the Australian researcher).

Although the six web documents contained different information and points of view about the same issue, students rarely referred to source information about documents other than the one they were reading at any given time. The mean amount of references to other source was 2.08 ($SD = 3.74$). The document from Today’s Medicine elicited the highest amount of intertextual sourcing ($M = .65$, $SD = 1.11$), with the majority (66.7%) referring to Illustrated Science. This is probably due to the fact that the claims of the Australian researcher were central in both documents, and that the documents took different stances regarding those claims.
We expected that students would pay attention to different source features while reading the different web documents. Table 1 shows that explicit references to present source focused on the author, the publication, the publisher, the date, and the type of document. The results indicate that students spontaneously pay attention to different source features for different documents. In particular, attention to author, publisher, and type of document seem to vary across the documents. Students paid attention to details about the author when they read Today’s Medicine more frequently than when reading any of the other documents. This was a debate article promoting a specific stance on the issue and students needed to note the author’s affiliation to infer why he might be biased, as commented by one student: “I forgot to read the information at the top. He is a chief engineer representing the Cell Phone Operators’ Association. I thought he (the author) was a medical doctor, since this is Today’s Medicine.” However, approximately half of the students did not note this, as for example: “And this is Today’s Medicine, independent trade journal for the health sector. Never heard of it.” By noting that it was an independent journal for the health sector, this student may trust the content more than students noting that it was a debate article.

A majority of the students noted the publisher of the NRPA document, whereas they did not for the other documents. This could be explained by the fact that this web document did not include any information about author or publication, and that many people know and trust the NRPA (Vinsand & Norsk Respons, 2007). Students probably had less knowledge about Today’s Medicine and Idaho Observer, as those publications are less visible in Norwegian media, and more students paid attention to information about what type of publications those were.

In total, 37 (72.5%) of the students explicitly referred to source information about the web documents (present source). When we look at sourcing activity related to each web document separately, however, no more than approxi-
Approximately half of the students explicitly referred to source information about any of the documents. We registered the highest number of students (28) explicitly noting source information about a web document (present source) for the NRPA document, while the lowest number (15) was registered for the textbook excerpt.

**Citations in essay**

In total 45 out of 51 students cited sources in their essays, with the mean number of citations being 4.58 ($SD = 3.35$). When we analyzed the difference between explicit and implicit citations for the six documents, a Wilcoxon Signed Ranks test indicated that the mean number of explicit citations ($M = 2.28$, $SD = 2.49$) were statistically significantly higher than implicit citations ($M = 1.25$, $SD = 1.49$), with $z = 2.18$, $p < .05$, $r = .22$. Students also included statistically significantly more citations to the six documents (present source: $M = 2.91$, $SD = 2.69$) than to sources embedded in the documents (embedded source: $M = 1.67$, $SD = 1.65$), with $z = 2.65$, $p < .01$, $r = .26$.

Of note is that the majority of the embedded source citations (82.4%) were presented as sources independent of the documents they were cited in. For example, one of the students referred to a disputable claim in the Idaho Observer, where the journalist argued that important research results concerning the danger of using cell phones have been denied publication in the two prestigious journals *Nature* and *Science* because the results could cause panic. The student seems to forget that the Idaho Observer is the primary source in this case, instead referring to all three publications as independent sources: “Several of the conclusions in the documents I have read seemed to be made up. I say this because the sources did not seem trustworthy, as for example the Idaho Observer, and the journals *Nature* and *Science.*” Thus, the student does not distinguish between the documents read and sources embedded in the documents.

Figure 3. Citations in essays specified by where the source information was located in the web documents.
Figure 3 shows the frequency of citations from the different web documents. A non-parametric omnibus Friedman test was performed, indicating statistically significant differences in the mean number of citations between documents, $\chi^2(6, N = 51) = 43.71, p < .001$. A follow-up Wilcoxon Signed Ranks test showed that the difference between the most cited web document (Illustrated Science) and the second most cited document (NRPA) was statistically significant, $z = 2.03, p < .05, r = .20$. However, we found that this difference was mainly due to the relatively high number of embedded source citations for the Illustrated Science ($M = 1.02, SD = 1.17$), while the number of present source citations was actually highest for the NRPA document ($M = .80, SD = 1.23$). The relatively high number of present source citations for the NRPA document may be related to the fact that this document, together with the textbook, was shown to be considered more trustworthy than the other documents by students in a prior study (Bråten et al., in press). One should, however, note that the textbook, also considered trustworthy, was cited less than any of the other documents in the essays ($M = .14, SD = .45$).

**Relationships between sourcing while reading and citations in essays**

We studied the relationship between sourcing while reading and citations in essays by computing non-parametric correlations (Kendall’s $\tau$). We did not include data for all separate documents in the analyses, but used sum variables. Our assumption was that the students’ total sourcing activity across the documents during reading might relate to citations in the essays, and we decided to focus on present source and total sourcing activity. Likewise, we included citations referring to present source in the essays, and here we focus on explicit present source citations since that is the norm in academic writing.

**Table 2. Nonparametric correlations between online sourcing and citations in essays**

<table>
<thead>
<tr>
<th>Variable</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Reading: Present source</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2. Reading: Explicit Present source</td>
<td>.82***</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3. Reading: Total sourcing</td>
<td></td>
<td>.65***</td>
<td>.60***</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4. Essay: Explicit Present source</td>
<td>.43***</td>
<td>.51***</td>
<td>.30**</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5. Essay: Total citations</td>
<td></td>
<td></td>
<td>.33***</td>
<td>.37***</td>
<td>.25*</td>
</tr>
</tbody>
</table>

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

Table 2 shows that student references to present source while reading correlated statistically significantly with explicit present source citations in essays ($\tau = .43, p < .001$), and also with the total number of citations ($\tau = .33, p < .001$). Focusing on only explicit spontaneous sourcing, the relationship to citations in essays becomes even stronger, both to explicit present source citations ($\tau = .51, p < .001$) and total number of citations ($\tau = .37, p < .001$). The
relationship between the total amount of sourcing while reading and the total number of citations in essays (τ = .25, p < .05) indicated that the inclusion of references to embedded sources did not contribute positively to the relationship. In sum, the results indicated, in accordance with our expectations, that students noting source information while reading multiple web documents also tend to include explicit references to those documents in essays.

DISCUSSION

Our purpose in the present study was to investigate student attention to and use of source information in reading and writing from multiple web documents. In the following, we discuss the results in relation to spontaneous sourcing while reading, citations in essays, and relations between spontaneous sourcing and citations in essays, respectively.

Students’ spontaneous sourcing while reading

With respect to spontaneous sourcing, the findings indicated that when reading multiple and partly conflicting web documents on a controversial topic, students paid attention to source information, although there was great variation in student sourcing activity. Moreover, the results revealed that students may not only pay attention to source information about documents, but also to source information embedded in the web documents. This is an important finding as many expository documents include source information at different levels, and the reader should be aware of the relationship among those levels. The present study demonstrates that when students read multiple expository web documents, they may also pay attention to different levels of source information and even confuse the difference between those levels by, for example, judging the trustworthiness of a web document and sources cited in the document on equal terms.

In addition, the results showed that students had the highest frequency of utterances related to source information about the documents when they read the two web documents representing the strongest contradictory claims. This is consistent with the findings of Braasch, Rouet, Vibert, and Britt (2012), indicating that inconsistent or contradictory information may make students more aware of source information, whereas documents presenting more neutral information may not create the same need to pay attention to the source of that information. Paying attention to source information about contradictory documents may represent an attempt to solve – or at least to understand and evaluate – the claims constituting the contradiction (Stadtler & Bromme, 2013).

We expected that students would, to some degree, also show signs of attention to source information about web documents they were not currently reading, that is, of intertextual sourcing. This could have indicated that students established relationships between documents, as described in the Inter-
text model (Britt & Rouet, 2012: Perfetti et al., 1999). Our results, however, indicated that undergraduates did not do that while reading the web documents. Although we found traces of such model construction, especially when students processed documents strongly contradicting each other, students’ overall spontaneous sourcing activity did not suggest they were in the process of constructing a full Intertext model of the document set. Rather, the results might imply that students at this level construct incomplete Intertext models by linking some selected documents but not all.

Finally, our results concerning spontaneous sourcing suggested that students, to some extent, may pay attention to different source features depending on the web documents. This is consistent with studies asking students to specify which source features they emphasize when judging documents’ trustworthiness after reading (Bråten et al., in press; Rouet et al., 1996). When venue or information about the publication is available this may be the feature most students note across documents. For two of the web documents in our set (Today’s Medicine and Idaho Observer), information about the author and the type of document also seemed to be important. This may be explained by the fact that both publications were probably unfamiliar to many students and that they therefore felt the need to know more about what kind of documents they were dealing with. Both documents also included strong statements, which may have generated a need to know more about the web documents.

Citations in essays
Approximately half of the references to sources were implicit, which may indicate that participants lacked skills regarding academic citation norms. This point is also highlighted by the fact that students frequently treated sources cited in the web documents as independent sources, although such sources were used to support arguments made by the authors of the documents. This result may reflect lack of awareness of how sources are used in documents and of the different levels of source information in expository documents among students.

Relations between online sourcing and sourcing in essays
Student sourcing activities while reading were related to their use of explicit citations in the essays. This result clearly indicates that sourcing skills are not solely a question of academic reading skills or about academic writing. Rather, sourcing skills seem to include attention to relevant source information as well as awareness of how to use that information when expressing oneself in writing. This result is in accordance with the Britt and Aglinskas (2002) study, where students who were prompted to pay attention to source information while reading also had more explicit references to sources in their essays. Attention to source information while reading may be a precondition to academic writing, and it seems to be essential that students pay attention to source information about the documents themselves and not only to source information embedded in the documents.
CONCLUSION

The Internet represents a wealth of information resources, offering opportunities for students to access information about nearly everything they may wish to study. Also, the Internet offers opportunities for everyone to publish, whether their message is reliable, adequate, honest, or not. As a consequence, students should be critical of the information they encounter on the Web (Leu, Kinzer, Coiro, Castek, & Henry, 2013). Paying attention to and reflecting on source information are essential parts of such critical reading.

Results from the present study show that undergraduates – to some extent – paid attention to source information, that they reflected on that information to a lesser degree, and that they had problems differentiating sources cited in a web document and the author of that document. Sourcing skills are an important aspect of academic reading and writing, and spontaneous sourcing while reading is related to students’ use of source information, that is, to their citations in written reports. Consequentially, teachers should not only focus on developing student citation skills in the context of writing tasks; student sourcing while reading should also be of concern to teachers. The study also indicates that undergraduate readers may not construct a full Intertext model (Perfetti et al., 1999; Britt & Rouet, 2012), that is, a representation of “who said what and how is that related to what other sources say”. Students at this level obviously need some guidance in becoming aware of the importance of source information and in sorting out the different layers of source information, with this being potentially profitable for text comprehension and academic citation skills alike.

The exploratory nature of the study implies a need for more research. More studies on how to facilitate the development of student sourcing and citation skills are much needed. Prior studies indicate that training sourcing skills may help in developing student attention to source information, citation skills, and text comprehension (Britt & Aglinskas, 2002; Britt et al., 2004; Graesser, Wiley, Goldman, O’Reilly, Jeon, & McDaniel, 2007). Those studies included specific digital tools to scaffold student sourcing activities. In addition, more attention should be directed towards initiatives seeking to incorporate training in sourcing and citation skills in even more ecologically valid educational settings, for example, in the form of discussions about texts and teacher feedback on assignments (see Braasch, Bråten, Strømsø, Anmarkrud, & Ferguson, 2013; Nokes, Dole, & Hacker, 2007; Reisman, 2012). We believe this to be central, not only because sourcing activities are important academic skills, but also because sourcing skills are important in daily life, where people search for and encounter an ever-increasing multitude of information sources of varying quality.
APPENDIX 1

CODING OF ONLINE SOURCING: CATEGORIES AND SAMPLE UTTERANCES

<table>
<thead>
<tr>
<th>Category</th>
<th>Examples</th>
</tr>
</thead>
<tbody>
<tr>
<td>Explicit sourcing</td>
<td>This is from the Illustrated Science.</td>
</tr>
<tr>
<td></td>
<td>This was published in 2008 in Dagbladet (newspaper)</td>
</tr>
<tr>
<td>Implicit sourcing</td>
<td>He may be exaggerating.</td>
</tr>
<tr>
<td></td>
<td>Look at this, wow, Idaho (incomplete)</td>
</tr>
<tr>
<td>Present source</td>
<td>The next one is a debate article in Today’s Medicine.</td>
</tr>
<tr>
<td></td>
<td>It’s an excerpt from a high school textbook.</td>
</tr>
<tr>
<td>Source in embedded source</td>
<td>This was claimed by Dr. Khurana.</td>
</tr>
<tr>
<td></td>
<td>I would like to check his website.</td>
</tr>
<tr>
<td>Other source in document set</td>
<td>This contradicts the prior article.</td>
</tr>
<tr>
<td></td>
<td>This was also stated in the first text, the excerpt from the textbook.</td>
</tr>
<tr>
<td>Attention</td>
<td>Let me see – it’s a research journalist</td>
</tr>
<tr>
<td></td>
<td>I’ll start by looking up the source. It’s a science textbook.</td>
</tr>
<tr>
<td>Evaluation</td>
<td>This seems to be important, the National Radiation Protection Agency.</td>
</tr>
<tr>
<td></td>
<td>Written by a chief engineer from the Cell Phone Operators’ Association – he may be biased.</td>
</tr>
</tbody>
</table>

Note

The protocols were coded three times according to: 1) whether statements were explicitly or implicitly referring to sources; 2) whether students were referring to source information about a document (present source), source information included in a document (source within embedded source), or source information from another document; and 3) the kind of activity the utterance indicated (attention and evaluating).

REFERENCES


