Using eye-tracking to assess sourcing during multiple document reading: A critical analysis

Ladislao Salmerón[a], Laura Gil [a], Ivar Bråten[b]

[A] Department of Developmental and Educational Psychology, University of Valencia, Spain
[B] Department of Education, University of Oslo, Norway

Abstract

During the last 15 years, there have been some efforts to extend the use of eye-tracking to researching reading in complex contexts, such as the reading of multiple documents. The research community involved in this extension has been interested in higher-order comprehension processes occurring in complex reading contexts, such as sourcing, defined as the processes of attending to, representing, evaluating, and using available or accessible information about the sources of textual content. In this article, we argue that extending eye-tracking research to investigate more complex reading contexts has been made without critically reflecting on its validity in those contexts. Specifically, because eye-tracking captures automatic as well as conscious processes, it is currently an open question how consistently eye-tracking captures the strategic sourcing processes that take place during multiple document reading, in particular when using real documents that include salient source information that may attract bottom-up fixations. In contrast, subjective methods, such as interviews, mainly target conscious processes, and may therefore be a more valid and generalizable measure of strategic sourcing activities. We compared sourcing indicators based on eye-tracking measures to sourcing indicated by a post-reading interview. Results suggested that current eye-tracking indices of sourcing are not universally valid measures, and that simpler methods, such as asking readers whether they paid attention to source information, may be more suited to assess strategic sourcing during multiple document reading.

Keywords: eye-tracking; interview; reading comprehension; multiple documents; sourcing.

1. Introduction

During the last four decades, reading researchers have extensively used eye-tracking measures to understand reading comprehension processes as they unfold during actual reading (for reviews, see Leinenger & Rayner, 2017; Rayner, 1998, 2009). This approach is based on the mind-eye hypothesis (Just & Carpenter, 1980), which assumes that readers fixate words as long as they are being processed. Therefore, eye-movements may be used as an indicator of what and how readers process when interacting with text. A major advantage of eye-tracking compared to subjective measures, such as verbal reports or interviews, is that this is an objective measure that captures subtle and automatic processes during reading, processes that may not be conscious to the reader. Our current understanding of reading processes owes much to the use of eye-tracking methodology, but the use of eye-tracking also comes with certain limitations. To be able to map specific eye-movements (e.g., regressions) onto specific reading processes (e.g., interpreting anaphoric references), researchers have developed highly controlled experimental stimuli. This approach has resulted in a research corpus on reading comprehension with an overrepresentation of simple text materials, such as sentences, paragraphs, or, at best, brief narratives (Jarodzka & Brand-Gruwel, 2017). However, during the last 15 years, there have also been some efforts to extend the use of eye-tracking to researching reading in more complex contexts, such as the reading of long expository texts (e.g., Catrysse, Gijbels, Donche, de Maeyer, Lesterhuis, & van den Bossche, 2018; Hyönä, Lorch Jr., & Kaakinen, 2002; Kaakinen, Hyönä, & Keenan, 2002), digital hypertexts (e.g., Salmerón, Baccino, Cañas, Madrid, & Fajardo, 2009; Salmerón, Naumann, García, & Fajardo, 2017), or literary pieces (e.g., Jacobs, 2016). The research community involved in this extension has been interested in higher-order comprehension processes occurring in complex reading contexts, such as the selection of information, identification of text relevance, and integration of textual information. In this article, we will argue that extending eye-tracking research to investigate more complex reading contexts has been made without critically reflecting on the validity of this approach when used in those contexts. Specifically, we will focus on the recent use of eye-tracking to study sourcing during the reading of multiple documents. Finally, we will present an empirical study to test the applicability of current eye-tracking indices to detect sourcing.

In the context of reading multiple documents, sourcing refers to processes of attending to, representing, evaluating, and using available or accessible information about the sources of textual content, for example about the author or publisher of different texts (Bråten & Braasch, 2018). When people read multiple documents about complex, controversial issues of which they have limited prior knowledge, such as climate change, they will have great difficulties determining the accuracy or truth value of information directly through "first-hand evaluation" (Stadtler & Bromme, 2014). In such complex reading contexts, they may therefore profitably resort to the strategy of evaluating information indirectly in light of the features of the sources (termed "second-hand evaluation" by Stadtler and Bromme). This is consistent with the Discrepancy-Induced Source Comprehension (D-ISC) model (Braasch & Bråten, 2017; Braasch, Rouet, Vibert, & Britt, 2012), which describes processes occurring when readers try to understand conflicting information presented by different sources. According to D-ISC, readers may find it difficult, if not impossible, to construct a coherent mental representation of an issue if conflicting information is presented across multiple sources. In such situations, readers will therefore strategically shift cognitive resources towards constructing a mental representation of the issue that includes source information (e.g., about the document authors) as organizational elements (Braasch & Bråten, 2017). In other words, conflicting information across multiple sources becomes an impetus for strategic sourcing

processes and such processes, in turn, allow readers to construct a meaningful interpretation of the issue despite existing conflicts. Because eye-tracking presumably captures automatic as well as conscious processes (McCormick, 1997), it is currently an open question how accurately and consistently eye-tracking captures the strategic sourcing processes described above, in particular compared to subjective methods that mainly target conscious processes, such as interviews.

In their recent review of the threats to validity in eye-movements research, Orquin and Holmqvist (2018) identified the distinction between bottom-up and top-down fixations as a major potential problem. Specifically, a stimulus can be fixated due to top-down processes, which indicates that readers strategically allocate their attention with the purpose of extracting information. For example, when reading a sentence, readers focus their attention sequentially on words to extract meaning. But a stimulus can also be fixated due to bottom-up processes, for example, because it is visually salient. Such fixations do not represent strategic actions, and are not implemented with the goal of extracting meaning from the stimulus. Bottom-up fixations may not be common while reading simple text materials typically used in laboratory settings, such as a paragraph written in uniform font size. This is because other information from the documents is eliminated to avoid distractions in such settings. Accordingly, the mind-eye hypothesis (Just & Carpenter, 1980) can accurately describe the knowledge extraction processes associated with reading in such settings, because most fixations are driven by top-down processes. However, bottom-up fixations may be frequent, and thus pose a threat to validity, when the texts are presented as real documents, for example as a real newspaper. In such a case, a fixation on the nameplate on the front-page of the newspaper (e.g. "The New York Times") may be triggered by readers' attempts to identify and evaluate the document source (i.e., top-down), or simply occur because the nameplate was a salient feature of the document (i.e., bottom-up). Therefore, interpretations of such fixations as purely strategic processes will be highly problematic.

During the last decade, several authors have attempted to use eye-tracking to measure sourcing in multiple document reading situations (Braasch et al., 2012; Brand-Gruwel, Kammerer, van Meeuwen, & van Gog, 2017; Gerjets, Kammerer, & Werner, 2011; Kammerer, Kalbfell, & Gerjets, 2016; Mason, Pluchino, & Ariasi, 2014; van Strien, Kammerer, Brand-Gruwel, & Boshuizen, 2016). These researchers have used different indices derived from fixations on source information available on pages, typically document logos (e.g., the name of a company), 'about us' information, or names of authors. Continuous indices include first and second pass fixation duration on source names in headlines or logos (Mason et al., 2014) and total fixation time on logos (Gerjets et al., 2011; van Strien et al., 2016). Non-continuous indices include percentage of page logos fixated for more than 100 msec (Brand-Gruwel et al., 2017).

One set of studies have used eye-tracking measures to test the hypothesis that readers tend to consider sources in trying to resolve inconsistencies across documents, in accordance with the D-ISC model (Braasch & Bråten, 2017; Braasch et al., 2012). For example, authors' affiliations could be used to interpret the extent to which they are knowledgeable about the issue and likely to provide unbiased information about it. Braasch et al. (2012, Exp 1) had participants read brief texts in which two authors provided either consistent or inconsistent accounts of the same event. When reading inconsistent accounts, participants were found to refixate the source area including information about author affiliation more often than when reading consistent accounts. More recently, Kammerer et al. (2016) replicated this effect in a multiple document scenario. In two studies, participants read two webpages about a controversial health-related issue that provided either consistent or inconsistent information about the issue. Eye-tracking data were used to measure the time spent on reading the 'about us' information when encountering consistent versus inconsistent information across webpages. Supporting the D-ISC model, participants spent more time reading the source information when reading inconsistent compared to consistent information.

However, another set of studies have used eye-tracking indices of sourcing in a more exploratory (rather than experimental) manner, that is, without a clear connection to theoretical models of sourcing, such as the D-ISC model (Braasch & Bråten, 2017). Thus, none of these studies investigated sourcing during specific reading episodes, such as when encountering conflicting information, but rather tried to capture sourcing through eye-tracking from beginning to end. Results reported in those studies provide examples that question the validity of eye-tracking measures to accurately assess readers' sourcing activities. For example, participants explicitly told to assess the trustworthiness of webpages have not been found to spend more time inspecting source information from the pages (i.e. page logos) than those not told to do so (Gerjets et al., 2011).

Moreover, a range of studies within multiple document literacy has indicated that readers with higher topic knowledge typically display more sourcing than do readers with lower topic knowledge (Braasch, Bråten, Strømsø, & Anmarkrud, 2014; Bråten, Strømsø, & Salmerón, 2011; Rouet, Britt, Mason, & Perfetti, 1996; Rouet, Favart, Britt, & Perfetti, 1997; Salmerón, Kammerer, & García-Carrión, 2013; Stang Lund, Bråten, Brandmo, Brante, & Strømsø, in press; Strømsø, Bråten, & Britt, 2010). These studies have used various methodologies, such as expert-novice comparisons or think aloud protocols, to capture the effects of prior topic knowledge on sourcing measured offline (e.g., by means of argumentative essays of source memory). However, attempts to replicate this effect using eye-tracking indicators of sourcing have been unsuccessful. For example, Brand-Gruwel et al. (2017) reported no difference between experts and novices in the percentage of webpages on which source information was fixated, and van Strien et al. (2016) found no difference between students with more or less prior topic knowledge in this regard.

Finally, researchers within multiple document literacy have identified a positive relationship between sourcing and the comprehension of multiple conflicting documents (Anmarkrud, Bråten, & Strømsø, 2014; Barzilai & Eshet-Alkalai, 2015; Barzilai, Tzadok, Eshet-Alkalai, 2015; Bråten, Strømsø, & Britt, 2009; Goldman, Braasch, Wiley, Graesser, & Brodowinska, 2012; Salmerón, Gil, & Bråten, 2018; Strømsø et al., 2010; Wiley et al., 2009). To the best of our knowledge, only one study has directly tested the relationship between sourcing measured via eye-tracking and students' performance. In that study, van Strien et al. (2016) found a weak correlation between total fixation times on logos and number of ideas included in a post-reading summary, but only for documents that provided information that was in line with participants' own attitudes towards the topic (but not for documents that contradicted participants' own attitudes towards the topic).

In sum, existing evidence suggests that current eye-tracking indices of sourcing are not universally valid measures, and that they may be especially problematic when they are not tied to experimental manipulations but, rather, used in exploratory studies. In the present study, we built on the D-ISC model in assuming that sourcing represents conscious processing in the service of constructing coherent mental representations (Braasch & Bråten, 2017), and that it, therefore, will be more adequately measured by means of a subjective measure, such as an interview, than by means of eye-tracking. Therefore, we expect that sourcing as measured by what could be considered gold standard measures (i.e., source citation in essays and memory for sources) would be predicted by participants' reporting of sourcing activities in a post-reading interview, but not by different eye-tracking indicators of attention to source information during reading.

2. Method

2.1. Participants

Forty-three undergraduates from a public university in eastern Spain participated for extra course credit. The sample included 33 females and 10 males, who ranged in age from 17 to 32 years and had an overall mean age of 19.9 (SD = 2.7). All participants were Caucasians and native Spanish speakers. Participants were assigned randomly to conditions in which they read either real or print-out versions of multiple documents (see sections 2.2.3 and 2.3 below).

2.2. Instruments and materials

2.2.1. Documents

Participants read four authentic documents containing relevant information about the topic of climate change. The first document was a 526-word excerpt from a textbook that provided basic knowledge about the greenhouse effect and described current controversies concerning climate change. The second document was a 363-word editorial from a major Spanish newspaper focused on the negative consequences of climate change. The third document was a 296-word blog entry written by an expert in climatology that discussed the potential benefits of increases in CO2. The last document was a 286-word article from a popular science magazine arguing that efforts to prevent global warming, although necessary, will have great economic costs. In addition to information about the sources of the documents themselves, all the documents cited at least two other sources (i.e., contained embedded sources) but none of the documents referred to each other. More specifically, the textbook excerpt included two embedded sources, while the newspaper editorial included four, the blog entry three, and the popular science article two embedded sources. The number of embedded sources was identical in the print-out and real document conditions. A more detailed description of these four documents can be found in Salmerón et al. (2018).

Documents were presented in two conditions that varied in their design: real documents or print-out copies of them (see below). Such manipulation was originally used to test the document as entities hypothesis (Britt, Rouet, & Braasch, 2013), which suggests that readers' integration of information across documents is likely to increase when source characteristics are more salient. In the context of the current manuscript, this manipulation served to compare the validity of eye-movements indices when reading complex real documents as compared to visually more uniform print-out versions, as well as to test the generalizability of the results to different types of texts.

In the real documents condition, participants read the textbook excerpt in a printed textbook, the newspaper editorial in a printed newspaper, the blog entry on a tablet, and the popular science article in a printed popular science magazine. Thus, in this condition, the documents differed with regard to a number of physical properties, such as size, weight, shape, and color, and participants had to adapt their body posture and their hands to read each text. When presented with the textbook, the newspaper, and the popular science magazine, participants were instructed to look for post-it notes that indicated the start and the end of the text to be read. In the print-out condition, participants read print-out versions of the same documents, with each text printed in black on a separate A4 sheet of paper, using Times New Roman, font size 12, and line space 1.5. Exactly the same source information that was available about the documents in the real document condition (i.e., author, document type, publisher, and date) was presented at the beginning of the texts, and the same embedded sources were included in the print-our versions of the documents.

2.2.2. Writing task

After reading the documents, all participants wrote an argument essay given the following written instruction: *The texts that you just read present different perspectives on climate change. Please write a report where you describe and evaluate the different perspectives, taking the arguments and evidence presented in the texts into consideration. Use approximately half a page for your essay.*

All essays were coded to indicate students' sourcing. In doing so, we identified all the segments that included specific references to source information from one or more of the documents as source citations. A segment was coded as a source citation when it referred to either accurate source feature information about the documents themselves, for example to document author, document type, or publication (e.g., "as described in the textbook", "the text from the newspaper explains", "the article published in the newspaper El País"), or to embedded sources (e.g., "at the end of the 19th century Arrhenius prophesied", "according to the IPCC").

Two independent judges independently scored a random selection of 10% of the essays. The Cohen's Kappa agreement was .84 for the coding of students' source citations, which indicates an almost perfect agreement according to Landis and Koch (1977). All disagreements were resolved in thorough discussion between the two raters, and one of them scored the remaining essays with respect to source citations using the same coding system.

2.2.3. Source memory task

Participants were requested to write down all the information they remembered about the documents they had just read, including the author, the document type, the publisher, and the date. For each document, we coded three source features: the author or publisher (i.e., the surname of the document's author or the name of the document's publisher), the document type (i.e., book or textbook/editorial, newspaper, or journalistic article/blog or webpage/magazine, science magazine, or science article), and the year of publication. The reason author and publisher were not scored separately is that the source information provided for the textbook excerpt and the newspaper editorial did not mention the author of the documents. Participants received one point for each source feature that they reported correctly (maximum score = 12 points). A random sample of 10% of the source memory statements were independently scored by two raters, yielding adequate reliability (Cohen's Kappa = .79). All disagreements were resolved in thorough discussion between the two raters, and one of them scored the remaining source memory statements according to the same coding system.

2.2.4. Interview

Following completion of the source memory task, the first or second author interviewed participants about their reading of the documents. During the interview, participants were asked questions about their use of specific strategies while reading the documents and their perceptions of the potential benefits of using each strategy. For the purpose of this study, only responses to a question concerning participants' attention to source information were analyzed. Specifically, they were asked: *While reading the texts, did you ever pay attention to the source of the document, such as the author or publisher ...?* If needed, the interviewer provided more examples to ensure that the student had understood the question correctly. Participants' responses to this question were scored 1 when they stated that they had paid attention to source information during reading and 0 when they said they had paid no attention to sources.

2.2.5. Apparatus

Students' eye-movements during the reading task were assessed by an eye-tracker mounted in a pair of glasses (SMI eye tracking glasses), with a sampling rate of 30 Hz. The selection of this apparatus, over more robust ones, was due to the fact that the stimuli were provided in paper, and therefore the glasses were the only available eye tracking equipment to use. The SMI eye tracking glasses use an event detection algorithm that identifies saccades and blinks, and classifies the remaining samples as visual intake or fixation. Fixations shorter than 50 msec are excluded. A pre-processing algorithm corrects for head movements.

In the print-out condition, the area of interest (AOI) corresponding to source information was a short text that described the type of document, the name of the publisher, the author's name, and the date. For example, the popular science magazine included the following source information: *Article from the journal Scientific American, by John Matson, journalist. Number 76, 2nd trimester 2014.* This text was always located in the upper right part of the print-out (Figure 1, left). In the real document condition, source information was located in different AOIs and included the logo of the publisher, the author's name, and the date (Figure 1, right). For example, source information for the popular science magazine included the name of the journal and the date on the front page, and the author's name and occupation below the title on the text page.

We computed two types of indices of sourcing based on eye-movement to the source AOIs: continuous and non-continuous. Continuous indices included a) first pass dwell time that started the first time students entered an AOI and ended when they shifted to a different AOI (in milliseconds), b) number of fixations of a first pass dwell episode, c) second pass dwell time corresponding to all subsequent revisits to that source's AOI, and d) number of fixations of all second pass dwell episodes. Non-continuous indices were based on the number of sources ($N = 4$) fixated for a particular time period. As there are no clear guidelines to determine a particular threshold, we explored four different time thresholds. Specifically, we computed the total number of sources fixated for at least 100, 500, and 1000 milliseconds (cf. Brand-Gruwel et al., 2017; Kammerer & Gerjets, 2014).

Bromme, Stadtler, and Scharrer (2018) have argued that source information is not only meta-textual, as some relevant aspects, such as document genre, can be derived from the text itself. For example, science magazines tend to use neutral voice, while blogs tend to be written in first-person voice. In an attempt to isolate potential effects of text processing on source identification, we included total text reading time and total number of text fixations as covariates in our analyses.

Figure 1. Examples of the newspaper editorial presented in the print-out (top) and real document (bottom) conditions. AOIs for the content are represented by solid line squares, while AOIs for source information are indicated by dash line squares. Note that in the real document condition the source information (nameplate and date) was located on the front-page and the content on an inside page.

2.3. Procedure

Participants were tested individually in sessions lasting approximately 90 minutes. On arrival, they were sequentially assigned to either the real document condition ($n = 20$) or the print-out condition ($n = 23$). First, they answered a questionnaire on demographics and completed prior knowledge and interest measures (not reported in this study, but see Salmerón et al., 2018). After having completed these measures, they were presented with the documents and orally given the following instruction: *You are now going to read four texts that present different perspectives on climate change. After reading the texts, you are going to write a report that discusses the different perspectives based on the arguments and evidence presented in the texts. There is no time limit, and you can read and reread the texts in the order you prefer. Please note that you will not have the texts available while writing your report.* The presentation order for the four documents was randomized for each participant in the real document condition, and each participant in the print-out condition was presented with the documents in the same random order as the preceding participant in the real document condition. In both conditions, the four documents were piled on a table in random order and all were available to participants during the entire reading task. Calibration was set by using an A4 sheet of paper with three points (two in the upper corners, and one in the middle of the bottom of the page) that students held as if they were reading it. After participants had finished reading the documents, they completed the writing task, a text comprehension measure (also not reported in this study), and the source memory task in that order on a laptop using a word-processing application. Finally, participants were interviewed about their strategy use when reading the documents.

3. Results

First, we describe the different sourcing indices that we explored by means of eye tracking and compare them across the two conditions (i.e., real vs. print-out versions of documents). Next, we analyze the extent to which sourcing based on both continuous and discontinuous eye-tracking data, as well as sourcing reported in the interview, were related to students' offline sourcing (i.e., source citation and source memory). Finally, we triangulate the different sourcing indices extracted from eye-tracking and interview data.

3.1. Descriptive analyses

Across conditions, sourcing as indicated by both continuous and non-continuous eye-tracking data took longer time and was more frequent in the print-out than in the real document condition (Table 1). This pattern may simply be explained by the fact that in the print-out version, the amount of source information was greater (as the type of document was made explicit) and more salient (located together rather than distributed across several pages). In contrast, the number of participants who said in the interview that they had attended to source information during reading was similar in the two conditions.

Table 1

Descriptive statistics for all the measures by condition (real document and print-out versions).

	Real documents		Print-out versions		Skewness (SE) Kurtosis (SE)
	Mean	*SD*	Mean	*SD*	
Memory for sources	4.33	.85	3.56	2.50	.41 (.36) 1.07 (.70)
Source citations	2.95	1.74	1.52	1.70	.65 (.36) -.65 (.70)
First pass dwell time on sources [1]	620.20	233.04	1099.48	963.70	1.32 (.36) 3.20 (.71)
Second pass dwell time on sources [1]	500.53	562.39	2279.59	3172.38	1.34 (.36) 1.64 (.71)
Number of first pass fixations on sources [1]	3.60	1.25	5.78	4.73	1.16 (.36) 3.17 (.71)
Number of second pass fixations on sources [1]	3.09	3.17	11.29	15.45	1.21 (.36) 141 (.71)
Total sources fixated for at least 100 msec	2.85	.91	3.23	.99	-.44 (.36) (-1.10) (.71)
Total sources fixated for at least 500 msec	1.71	.95	2.80	1.28	.02 (.36) (-1.26) (.71)
Total sources fixated for at least 1000 msec	1.00	.94	2.28	1.48	.40 (.36) (-1.08) (.71)
Participant reportedly looked at sources while reading	12 / 21		13 / 22		

[1] Squared values were used to correct the positive skewness.

3.2. Sourcing as measured by continuous eye-movement indices

To test the validity of the continuous eye-movement indices of sourcing, we performed partial correlations comparing those values with offline sourcing as indicated by source citations in the essays and source memory, after controlling for the effect of total text reading time (for the indices based on dwell time) and total number of fixations on the text (for the indices based on fixation number). In the real documents condition, there were no significant correlations between the eye-tracking measures and indicators of offline sourcing (Table 2, first four columns). In the print-out condition, however, second-pass measures correlated positively with source citations, but not with source memory (Table 2, last four columns).

Table 2

Partial correlations between sourcing as measured by continuous eye-movement indices and offline indicators of sourcing by condition (real documents and print-out versions), controlling for the effect of total text reading time (for the indices based on dwell time) or total number of fixations on the text (for the indices based on fixation number).

	Real documents				Print-out versions			
	First pass dwell time on sources	Second pass dwell time on sources	Number of first pass fixations on sources	Number of second pass fixations on sources	First pass dwell time on sources	Second pass dwell time on sources	Number of first pass fixations on sources	Number of second pass fixations on sources
Memory for sources	-.29	.04	-.33	-.01	.25	.30	.26	.27
Source citations	-.00	-.01	.14	-.01	-.01	.45*	.08	.44*

Note: * $p < .05$

3.3. Sourcing as measured by discontinuous eye-movement indices

We first compared the number of sources fixated by participants (maximum 4 sources), as a function of fixation threshold, by condition (real documents and print-out versions). The number of sources fixated varied considerable between thresholds, and across conditions (Table 3). Chi square analyses comparing condition indicated that there was no significant difference for the 100 msec threshold, $\chi^2(3)= 3.93$, $p = .27$, but there were significant differences for the 500 and 1000 msec thresholds ($\chi^2(4)= 13.47$, $p < .01$ and $\chi^2(4)= 12.44$, $p = .01$, respectively). For the higher thresholds, participants in the real document conditions fixated less sources than in the print-out version.

Table 3

Number of sources fixated by participants, as a function of fixation threshold, by condition (real documents and print-out versions) (in percentage of participants).

	Real documents					Print-out versions				
	0	1	2	3	4	0	1	2	3	4
100 msec	0%	4.8%	33.3%	33.3%	28.6%	0%	4.8%	23.8%	14.3%	57.1%
500 msec	4.8%	42.9%	33.3%	14.3%	4.8%	4.8%	19%	4.8%	33.3%	38.1%
1000 msec	33.3%	42.9%	14.3%	9.5%	0%	19%	9.5%	23.8%	19%	28.6%

Next, to test the validity of the discontinuous eye-movement indices of sourcing, we computed partial correlations between the discontinuous indices of sourcing and the offline sourcing measures, controlling for the effect of total text reading time. In the real documents condition, there were no correlations between the eye-tracking measures and offline sourcing (Table 4, left columns). In the print-out condition, however, there were positive relations between number of sources fixated for at least 100, 500, and 1000 msec and source citations (Table 4, right columns). Memory for sources correlated only with the index using the 100 msec threshold.

Table 4

Partial correlations between sourcing as measured by discontinuous eye-movements indices and offline indicators of sourcing by condition (real documents and print-out versions), controlling for the effect of total text reading time.

	Real documents			Print-out versions		
	Total sources fixated for at least 100 msec	Total sources fixated for at least 500 msec	Total sources fixated for at least 1000 msec	Total sources fixated for at least 100 msec	Total sources fixated for at least 500 msec	Total sources fixated for at least 1000 msec
Memory for sources	.01	.06	-.06	.46*	.39	.22
Source citations	-.17	-.17	-.09	.49*	.52*	.51*

Note: * $p < .05$

3.4. Sourcing as measured in the interview

To assess the predictive power of sourcing as measured in the interview, we computed a series of ANOVAs with condition (real vs. print-out versions of documents) and reported sourcing (yes or no) as independent variables, and the two offline sourcing measures as dependent variables.

There were main effects of reported sourcing on the offline sourcing measures, with students reportedly having attended to sources while reading scoring higher on the offline measures (see Table 5, right column). None of the interactions between condition and reported sourcing were significant. In other words, the effects of reported sourcing were independent of whether students read real or print-out versions of the documents.

Table 5

ANOVAs of reported sourcing (as measured in the interview) and condition (real documents and print-out versions) on offline indicators of sourcing.

	Real documents		Print-out versions		
	Participant reportedly looked at the sources	Participant reportedly didn't look at the sources	Participant reportedly looked at the sources	Participant reportedly didn't look at the sources	Effect of sourcing
N	12	9	13	9	-
Memory for sources	4.67 (.89)	3.89 (.30)	4.31 (2.84)	2.56 (1.74)	$F(1, 39)= 4.92$, $p = .03$, $\eta^2_p =$ .11
Source citations	3.42 (1.88)	2.33 (1.41)	2.08 (2.01)	.89 (.78)	$F(1, 39)= 4.84$, $p = .03$, $\eta^2_p =$ .11

3.5. Triangulation of the sourcing measures

Finally, we compared sourcing as indicated by the different eye-movement indices and the interview. Specifically, we ran a set of ANOVAs with reported sourcing as independent variable and eye-movement indices as dependent variables (see Table 6).

Overall, there were no major differences in sourcing as indicated by the eye-movements between students who reportedly had or had not paid attention to sources while reading. Thus, regardless of what they reported about sourcing in the interview, students apparently looked at source information to some extent. The only significant difference emerged for the number of first pass fixations on source information, which increased for students who reportedly paid attention to sources while reading. This effect was especially pronounced in the print-out versions of the documents (Table 6, right column). In this condition, the trend was observed for other eye-tracking indicators, with participants who reportedly paid attention to sources seemingly looking longer at and fixating more on source information. However, such differences did not reach statistical significance, probably due to a lack of statistical power.

Table 6

ANOVAs of reported sourcing (as measured in the interview) and condition (real documents and print-out versions) on two sets of dependent variables (continuous and discontinuous eye-movement indices of sourcing).

	Real documents		Print-out versions			Interaction sourcing (interview) and condition
	Participant reportedly looked at the sources	Participant reportedly didn't look at the sources	Participant reportedly looked at the sources	Participant reportedly didn't look at the sources	Effect of sourcing (interview)	
First pass dwell time on sources	619.27 (232.18)	621.44 (248.29)	1402.89 (1061.67)	606.72 (378.28)	$F(1, 39) = 3.99, p = .06, \eta^2_p = .09$	$F(1, 39) = 4.06, p = .051, \eta^2_p = .09$
Second pass dwell time on sources	508.08 (605.22)	490.47 (535.59)	2986.39 (3790.17)	1582.19 (2182.04)	$F(1, 39) = 1.58, p = .22, \eta^2_p = .04$	$F(1, 39) = 1.13, p = .29, \eta^2_p = .02$
First pass fixations on sources	3.81 (1.48)	3.33 (.87)	7.62 (5.29)	2.83 (1.74)	$F(1, 39) = 9.25, p = .004, \eta^2_p = .20$	$F(1, 39) = 6.13, p = .01, \eta^2_p = .14$
Second pass fixations on sources	3.08 (3.33)	3.11 (3.14)	15.23 (18.14)	7.22 (10.88)	$F(1, 39) = 2.80, p = .16, \eta^2_p = .05$	$F(1, 39) = 1.74, p = .19, \eta^2_p = .04$
Sources fixated at least 100 msec	2.92 (.79)	2.78 (1.09)	3.64 (.67)	2.89 (1.17)	$F(1, 39) = 2.30, p = .14, \eta^2_p = .06$	$F(1, 39) = 1.08, p = .30, \eta^2_p = .03$
Sources fixated at least 500 msec	1.75 (.96)	1.67 (1.00)	3.36 (.92)	2.22 (1.48)	$F(1, 39) = 3.15, p = .08, \eta^2_p = .08.$	$F(1, 39) = 2.35, p = .13, \eta^2_p = .06$
Sources fixated at least 1000 msec	1.00 (.95)	1.00 (1.00)	2.81 (1.25)	1.67 (1.65)	$F(1, 39) = 2.22, p = .14, \eta^2_p = .06$	$F(1, 39) = 2.22, p = .14, \eta^2_p = .05$

4. Discussion and conclusions

Results from our study converge with existing evidence suggesting that current eye-tracking indices of sourcing have several limitations (Brand-Gruwel et al., 2017; Gerjets et al., 2011; Mason et al., 2014; van Strien et al., 2016), especially when they are not tied to experimental manipulations

performed to test specific hypotheses based on a theoretical model of sourcing, such as the D-ISC (Braasch & Bråten, 2017). Our results also suggest that simpler methods, such as asking readers whether they paid attention to source information, actually may be better suited to assess strategic sourcing during multiple document reading.

Specifically, sourcing measured by means of the interview predicted two different sourcing measures, source citations and source memory, independently of the type of reading materials (real documents or print-out versions). Relationships between sourcing measured by eye-movement indices and offline sourcing were more complex, however, as they depended on the type of reading materials and the type of indices used. Thus, neither continuous nor discontinuous eye movement indices predicted offline sourcing when reading real documents. This probably reflects the fact that fixations to source information in real documents is a combination of bottom-up and top-down processes, which makes such indicators less valid (Orquin & Holmqvist, 2018). Yet, for print-out versions of the documents, late eye movement continuous indices (i.e., second dwell time and number of second pass fixations), which usually reflect more strategic processing (Rayner, 2009), correlated positively with source citations, but not source memory. When using dichotomous eye movement indices sourcing as indicated by a 100 msec threshold positively correlated with both source memory and source citations, and higher thresholds (i.e., 500 and 1000 msec) positively correlated with source citations. Of note is that in these analyses, potential effects due to the time participants devoted to reading the actual text were controlled for.

The fact that participants accurately reported their sourcing activities in the interview support the idea that attending to and processing source information while reading multiple documents represent strategic and conscious activities, which is consistent with the D-ISC model recently elaborated by Braasch and Bråten (2017). Presumably, readers resort to this strategy in an effort to create coherent, meaningful mental representations from diverse documents on the same topic, for which they may need to include source information as organizational elements (e.g., to qualify claims according to source or better understand conflicts among sources) (see also van den Broek & Kendeou, 2015). As students are aware of such strategic sourcing activities in the service of meaning-making, post-task interviews seem suited to capture the use of sourcing in multiple document contexts. In comparison, measures based on eye-movements may incorporate a mixture of automatic and strategic processes, which make such indices less valid when used during the reading of more complex and diverse reading materials.

The mixture of automatic and strategic processes in eye-tracking data seemed to be particularly the case when participants read real documents, which included more visually salient information not present in the print-out versions. For example, the nameplate of the newspaper "El País" occupied a big portion of the upper part of the front page of the real document, while this source feature in the print-out version had the similar size as other textual information.

Overall, the results challenge the current use of eye-tracking indices solely based on fixations at source information (e.g., logos or 'about us' information). We suggest that future work could increase the validity of such measures in at least two different ways. First, although most students directly attend to source features for short periods of time, they may also evaluate sources by focusing on the quality of the arguments in the texts. That is, readers will not necessarily have to look at source information to judge the credible of a document. Accordingly, a combination of indices of source and text processing may provide a more valid picture of students' sourcing activities. Second, as indicated by our literature review, another way to improve the validity of eye-movement indices of sourcing in such challenging reading task contexts is to identify specific reading episodes in which sourcing may be expected based on theoretical assumptions, such as those forwarded by the D-ISC model (Braasch et al., 2012; Kammerer et al., 2016). Sourcing could

then be analyzed after the critical episode takes place, for example, after students finish reading a conflicting claim likely to involve a break in situational coherence. Presumably, this will reduce the probability that non-strategic, automatic processing is captured by the eye-tracking indicators. In the absence of such theoretical assumptions, when the goal of the study is to explore sourcing from beginning to end during reading, subjective methods such as interviews seem preferable when assessing strategic sourcing.

This study does not come without limitations, of course. Among them are the relatively small sample size, which made statistical power lower than desirable, the particular eye-tracker that we used, which may have influenced the accuracy level of the data, and the complexity of the reading materials that we presented. Moreover, it cannot be entirely ruled out that participants' reports of sourcing were influenced by their performance on the preceding source memory task. We could go on describing further potential limitations and qualifications. At the same time, given the sample, the apparatus, the reading materials, and the operationalizations that we did use, we believe our results still merit attention and hope they will provide impetus to more careful consideration of eye movement indices as valid measures of sourcing in future research.

Keypoints

- Use of eye-movements to investigate complex reading requires critical reflection on the validity of the measures
- Eye-movement indicators of students' sourcing do not always predict offline sourcing
- Interview data seem to be a more valid indicator of strategic sourcing in multiple document contexts

Acknowledgments

References

Anmarkrud, Ø., Bråten, I., & Strømsø, H.I. (2014). Multiple-documents literacy: Strategic processing, source awareness, and argumentation when reading multiple conflicting documents. *Learning and Individual Differences, 30,* 64-76.

Barzilai, S., & Eshet-Alkalai, Y. (2015). The role of epistemic perspectives in comprehension of multiple author viewpoints. *Learning and Instruction, 36,* 86-103.

Barzilai, S., Tzadok, E., & Eshet-Alkalai, Y. (2015). Sourcing while reading divergent expert accounts: Pathways from views of knowing to written argumentation. *Instructional Science, 43,* 737-766.

Brand-Gruwel, S., Kammerer, Y., van Meeuwen, L., & van Gog, T. (2017). Source evaluation of domain experts and novices during Web search. *Journal of Computer Assisted Learning, 33,* 234-251.

Braasch, J.L.G., & Bråten, I. (2017). The Discrepancy-Induced Source Comprehension (D-ISC) model: Basic assumptions and preliminary evidence. *Educational Psychologist, 52*, 167-

181.

Braasch, J.L.G., Bråten, I., Strømsø, H.I., & Anmarkrud, Ø. (2014). Incremental theories of intelligence predict multiple document comprehension. *Learning and Individual Differences, 31,* 11-20.

Braasch, J. L.G., Rouet, J. F., Vibert, N., & Britt, M. A. (2012). Readers' use of source information in text comprehension. *Memory & Cognition, 40*, 450-465.

Bråten, I., & Braasch, J.L.G. (2018). The role of conflict in multiple source use. In J.L.G. Braasch, I. Bråten, & M.T. McCrudden, M.T. (Eds.), *Handbook of multiple source use* (pp. 184-201). New York: Routledge.

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

Bråten, I., Strømsø, H.I., & Salmerón, L. (2011). Trust and mistrust when students read multiple information sources about climate change. *Learning and Instruction, 21,* 180-192.

Britt, M. A., Rouet, J. F., & Braasch, J. L. G. (2013). Documents as entities: Extending the situation model theory of comprehension. In M. A. Britt, S. R. Goldman, & J. F. Rouet (Eds.), *Reading: From words to multiple texts (pp. 160–179).* New York: Routledge.

Bromme, R., Stadtler, M., & Scharrer, L. (2018). The provenance of certainty: Multiple source use and the public engagement with science. In J. L. G. Braasch, I. Bråten, & M. T. McCrudden (Eds.), *Handbook of multiple source use* (pp. 269-284). New York, NY: Routledge.

Catrysse, L., Gijbels, D., Donche, V., de Maeyer, S., Lesterhuis, M., & van den Bossche, P. (2018). How are learning strategies reflected in the eyes? Combining results from self-reports and eye-tracking. *British Journal of Educational Psychology, 88*, 118-137.

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

Goldman, S.R., Braasch, J.L.G., Wiley, J., Graesser, A.C., & Brodowinska, K. (2012). Comprehending and learning from Internet sources: Processing patterns of better and poorer learners. *Reading Research Quarterly, 47,* 356-381.

Hyönä, J., Lorch Jr., R. F., & Kaakinen, J. K. (2002). Individual differences in reading to summarize expository text: Evidence from eye fixation patterns. *Journal of Educational Psychology, 94,* 44-55.

Jacobs, A. M. (2016). The scientific study of literary experience. *Scientific Study of Literature*, *5*, 139-170.

Jarodzka, H., & Brand-Gruwel, S. (2017). Tracking the reading eye: Towards a model of real-world reading. *Journal of Computer Assisted Learning, 33,* 193-201.

Just, M. A., & Carpenter, P. A. (1980). A theory of reading: From eye fixations to comprehension. *Psychological Review, 87,* 329-354.

Kaakinen, J. K., Hyönä, J., & Keenan, J. M. (2002). Perspective effects on online text processing. *Discourse Processes, 33,* 159-173.

Kammerer, Y., & Gerjets, P. (2014). The role of search result position and source trustworthiness in the selection of Web search results when using a list or a grid interface. *International Journal of Human Computer Interaction, 30,* 177–191.

Kammerer, Y., Kalbfell, E., & Gerjets, P. (2016). Is this information source commercially biased? How contradictions between web pages stimulate the consideration of source information. *Discourse Processes, 53,* 430-456.

Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics, 33*, 159-174.

Leinenger, M., & Rayner, K. (2017). What we know about skilled, beginning, and older readers from monitoring their eye movements. J. A. León & I. Escudero (Eds.), *Reading comprehension in educational settings* (pp. 1-17). Amsterdam: John Benjamins.

Mason, L., Pluchino, P., & Ariasi, N. (2014). Reading information about a scientific phenomenon on webpages varying for reliability: An eye-movement analysis. *Educational Technology Research and Development, 62,* 663-685.

McCormick, P. A. (1997). Orienting attention without awareness. *Journal of Experimental Psychology: Human Perception and Performance, 23,* 168-180.

Orquin, J. L., & Holmqvist, K. (2018). Threats to the validity of eye-movement research in psychology. *Behavior Research Methods, 50,* 1645-1656.

Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research. *Psychological Bulletin, 124,* 372-422.

Rayner, K. (2009). The 35th Sir Frederick Bartlett Lecture: Eye movements and attention in reading, scene perception, and visual search. *Quarterly Journal of Experimental Psychology, 62,* 1457-1506.

Rouet, J.F., Britt, M.A., Mason, R.A., & Perfetti, C.A. (1996). Using multiple sources of evidence to reason about history. *Journal of Educational Psychology, 88,* 478–493.

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

Salmerón, L., Baccino, T., Cañas, J.J., Madrid, R. I., & Fajardo, I. (2009). Do graphical overviews facilitate or hinder comprehension in hypertext? *Computers & Education, 53,* 1308-1319.

Salmerón, L., Gil, L., & Bråten, I. (2018). Effects of reading real versus print-out versions of multiple documents on students' sourcing and integrated understanding. *Contemporary Educational Psychology, 52*, 25-35.

Salmerón, L., Kammerer, Y., & García-Carrión, P. (2013). Searching the Web for conflicting topics: Page and user factors. *Computers in Human Behavior, 29,* 2161–2171.

Salmerón, L., Naumann, J., García, V., & Fajardo, I. (2017). Scanning and deep processing of information in hypertext: An eye-tracking and cued retrospective think-aloud study. *Journal of Computer Assisted Learning, 33*, 222–233.

Stadtler, M., & Bromme, R. (2014). The content-source integration model: A taxonomic description of how readers comprehend conflicting scientific information. In D.N. Rapp & J.L.G. Braasch (Eds.), *Processing inaccurate information: Theoretical and applied perspectives from cognitive science and the educational sciences* (pp. 379-402). Cambridge, MA: The MIT Press.

Stang Lund, E., Bråten, I., Brandmo, C., Brante, E.W., & Strømsø, H.I. (in press). Direct and indirect effects of textual and individual factors on source-content integration when reading about a socio-scientific issue. *Reading and Writing: An Interdisciplinary Journal.*

Strømsø, H.I., Bråten, I., & Britt, M.A. (2010). Reading multiple texts about climate change: The relationship between memory for sources and text comprehension. *Learning and Instruction, 20,* 192-204.

van den Broek, P., & Kendeou, P. (2015). Building coherence in web-based and other non-traditional reading environments: Cognitive opportunities and challenges. In R.J. Spiro, M. DeSchryver, M.S. Hagerman, P.M. Morsink, & P. Thompson (Eds.), *Reading at a crossroads? Disjunctures and continuities in current conceptions and practices* (pp. 104-114). New York: Routledge.

van Strien, J. L., Kammerer, Y., Brand-Gruwel, S., & Boshuizen, H. P. (2016). How attitude strength biases information processing and evaluation on the web. *Computers in Human Behavior, 60,* 245-252.

Wiley, J., Goldman, S.R., Graesser, A.C., Sanchez, C.A., Ash, I.K., & Hemmerich, J.A. (2009). Source evaluation, comprehension, and learning in Internet science inquiry tasks. *American Educational Research Journal, 46,* 1060-1106.