Using Eye-Movement Modeling Examples to Improve Critical Reading of Multiple Webpages on a Conflicting Topic

Ladislao Salmerón and Pablo Delgado

University of Valencia, Spain

Lucia Mason

University of Padova, Italy

May 2020: Accepted for publication at *Journal of Computer Assisted Learning*.

Correspondence about this manuscript can be sent to:

Ladislao Salmerón

Associate Professor of Educational Psychology

University of Valencia, Spain

Avd. Blasco Ibáñez, 21. 46010 Valencia, Spain

Ladislao.Salmeron@valencia.edu

Abstract

Interventions to promote students' source evaluations have used various methods designed for the classroom context. In the present study, we tested an approach that is easily adaptable to online courses, based on Eye Movement Modeling Examples (EMME), that is, short videos displaying an expert student's eye gaze while s/he reads multiple pages on the Internet to learn about a conflicting topic. Using an eye-tracking methodology in a pre-post design, we analyzed how EMME changed students' attention to source information, and how this processing affected their learning. EMME increased participants' attention to the search engine results page (SERP), author information, and decreased attention of texts from untrustworthy pages. In addition, EMME increased the number of participants who cited at least one document source at post-test. Finally, we discuss the potential benefits and limitations of EMME in teaching complex literacy strategies, and the importance of measuring processing data in educational research studies.

Keywords: eye-movement modeling examples**,** sourcing, eye-tracking, multiple-documents comprehension.

Using Eye-Movement Modeling Examples to Improve Critical Reading of Multiple Webpages on a Conflicting Topic

Introduction

Digital technology has produced an increase in written communication. For years, the Internet has been the most widely used source of information. Teenagers and adults of all ages spend an increasing amount of time reading texts on the Internet for various reasons. The large number of easily available texts puts great pressure on the individual's reading skills, well beyond simple word decoding and literal comprehension. Due to the lack of information "gatekeeping" on the Internet, users can often access misinformation or biased information, which can spread quickly and become "viral". Thus, more than ever, in the so called "post-truth" era of "alternative facts", critical reading becomes crucial (Bråten, Braasch, & Salmerón, 2019).

First, critical readers must pay attention to source characteristics, such as the author's credentials or the institution publishing the document. Second, they must evaluate the quality of the information by judging to what extent these source characteristics suggest that the information is supported by sound evidence, or on the contrary, could potentially be biased. Finally, critical readers must consider the conclusions from their assessment of the source to qualify the information accessed, e.g. by judging the view supported by expert sources as credible, or by discarding information from potentially biased sources (Tarchi & Mason, 2019). Because this is a sequential process, failing to attend to source characteristics may limit students' ability to critically read the information. Accordingly, educational interventions to improve critical reading tend to emphasize the importance of paying attention to source characteristics (Brante, & Strømsø, 2017). In their recent systematic review, Brante &

Strømsø (2017) identified two relevant gaps in the literature of source evaluation interventions. First, few intervention studies were designed specifically to guide students to critically read on the Internet. As this scenario presents different source features than traditional paper documents, such as search engine result page (SERP), participants may need specific guidance to find and use such source features. Second, none of the existing interventions used eye-tracking measures to identify how students look for source features or specifically which features they attend to during reading. As such, except for two studies that used think aloud protocols (Brand-Gruwel, & Wopereis, 2006) and navigation data (Stadtler, Paul, Globoschütz, & Bromme, 2015), the effectiveness of current interventions mostly relies on participants' off-line responses.

The present study tries to fill in those gaps by testing Eye Movement Modeling Examples (EMME) to support students' source evaluations on the Internet, while measuring changes in their visual attention to source characteristics from pre to post-test. EMMEs are videos that display, by means of a moving dot, where an expert is looking while performing an activity. By visually modeling expert behavior, EMMEs are intended to guide students' visual attention while performing a particular task (Jarodzka, van Gog, Dorr, Scheiter, & Gerjets, 2013; van Gog, Jarodzka, Scheiter, Gerjets, & Paas, 2009). The present study used this innovative approach to address one of the crucial questions in contemporary research on learning and instruction, that is, how to foster students' ability to critically evaluate information sources when reading webpages, starting from paying attention to source information.

Sourcing and Critical Reading on the Internet

In a seminal paper by Wineburg (1991), sourcing was defined as the process of using information about documents – author, genre, and publication date – to evaluate and interpret documents' content. Research has widely indicated that students do not tend to pay attention to source information, and this effect is generalized across different educational levels, including elementary school students (Kuiper, Wolman, & Terwel, 2008; Macedo-Rouet et al., 2013; Paul, Cerdán, Rouet, & Stadtler, 2018; Paul, Stadtler, & Bromme, 2019), middle school students (Mason, Ariasi, & Boldrin, 2011; Salmerón, Macedo-Rouet & Rouet, 2016; Salmerón, Agnese & Delgado, 2019; VanSledright & Kelly, 1998), high school students (Barzilai & Zohar, 2012; Macedo-Rouet et al., 2019; Paul, Macedo-Rouet, Rouet, & Stadtler, 2017; Walraven, Brand-Gruwel, & Boshuizen, 2009), and university students (Barzilai, Tzadok, & Eshet-Alkalai, 2015; Bråten, Strømsø, & Andreassen, 2016; Britt & Aglinskas, 2002; List, Du, Wang, & Lee, 2019; Salmerón, Gil, & Bråten, 2018b).

Undoubtedly, sourcing is an important part of critical reading in today's information society. Readers who are not able to identify and use source information to distinguish between trustworthy and untrustworthy sources also experience more difficulties in comprehending multiple conflicting texts on the same topic (e.g., Bråten, Ferguson, Strømsø, & Anmarkrud, 2014). In fact, sourcing skills are crucial not only to be able to follow reliable sources and rule out biased ones, but also to form an integrated and coherent representation of various texts despite their divergences. Research has indicated that memory for sources is associated with the construction of coherent representations of documents' contents (Salmerón et al., 2018b; Strømsø, Bråten, & Britt, 2010).

According to the seminal documents model framework (Perfetti, Rouet, & Britt, 1999; Rouet & Britt, 2011), the mental representation of a set of texts requires both the comprehension of each single text, that is, the situational model of each single text based on the integration of textual information and reader's prior knowledge (Kintsch, 1989), and two additional text-representation layers, the intertext and the integrated mental models of the various situations. The intertext model is formed when readers "tag" content to source characteristics such as author and genre. It refers to a representation of the sources of the multiple documents and their interrelationships.

The importance of tagging content and source features also emerged in the recent Discrepancy-Induced Source Comprehension (D-ISC) model (Braasch & Bråten, 2017). According to this model, readers pay more attention to "metadata" (who says what) when they are confronted with contrasting information on the same topic provided by different sources. The representation of source features also promotes the interpretation of documents' contents, contributing to the construction of an integrated mental representation of multiple documents. For example, by identifying that two authors have diverging motivations (e.g. to inform vs. to sell a product), readers can reconcile conflicting accounts about a controversial issue.

On the Internet, source features are available not only in the webpages (e.g. authors' credentials, logo of the Institution that hosts the webpage), but also in SERPs. SERPs display the title of the pages, a brief summary, and the URL, usually including the source that hosts the webpage. As such, they represent a unique opportunity for students to reflect on the relevance and quality of the pages and directly compare pages. In their recent study, Wineburg & McGrew (2017) emphasized the importance of attending to SERPs in learning about conflicting topics. In their study, professional fact checkers tend to spend some time scanning the SERP before clicking on any link, a strategy that the authors coined as "click restraint". They inspected the SERP to gather source information from different pages, and used this in their subsequent navigation. On the contrary, the group of undergraduate students tended to click on the first links of the list without spending much time on the SERP (for similar results see Brand-Gruwel, Kammerer, van Meeuwen, & van Gog, 2017). The importance of SERPs has also been identified in studies that manipulated the design to support source evaluation. For example, adapting the display of snippets to facilitate comparisons of webpages, such as presenting groups of snippets instead of a list, increased the probability of students including more trustworthy information in their essays (Kammerer & Gerjets, 2014) and receiving higher scores on an inter-textual comprehension measure (Salmerón et al., 2010). Thus, interventions aimed to improve critical evaluation of information sources on the Internet could benefit from directly addressing how to process SERP snippets.

To advance our knowledge about effective methods for promoting sourcing in critical reading, in the current study we adopted the approach based on EMME, aimed at modeling several strategic sourcing behaviors on the Internet: from carefully scrutinizing a SERP before selecting a trustworthy source to quickly skimming a visibly untrusworthy webpage.

Eye Movement Modeling Example

EMME combines the efficacy of video-based modeling and example-based instruction (e.g., Renkl, 2014) with the usefulness of eye-tracking methodology (e.g., Jarodzka & Brand-Gruwel, 2017). Rooted in observational learning (Bandura, 1977), video-based modeling has received increasing attention in educational research as a way to foster students' performance in diverse domains, such as making collages and writing poetry (Groenendijk, Janssen, Rijlaarsdam, & van den Bergh, 2013), solving problems about electric circuits (Hoogerheide, Renkl, Fiorella, Paas, & van Gog, 2019), or solving information problems when searching the Web for information (Frerejean, van Strien, Kirschner, & Brand-Gruwel, 2018). Examples are included in video-based modeling to help students focus on an appropriate way to execute a task (Hefter, ten Hagen, Krense, Berthold, & Renkl, 2019). Thus, they should not try to perform a task by themselves, but rather they can invest mental effort in achieving optimal performance based on the example, assuming that it is well-designed.

Eye-tracking technology provides videos of gaze replays. Therefore, it is not only useful for offering unique data about perceptual and cognitive processes, but also for providing the opportunity to show recorded visual behavior in the form of a video where fixations on specific information are represented as solid dots.

In the extant literature, there is now ample evidence of the effectiveness of EMME in various areas of investigation, such as classification tasks (Jarodzka et al., 2012), problem solving (van Marlen, Wermeskerken, Jarodzka, & van Gog, 2018), medical imaging (Litchfield, Ball, Donovan, Manning, & Crawford, 2010), medical education (Jarodzka et al., 2012; Seppänen & Gegenfurtner, 2017), digital reading (Salmerón & Llorens, 2019), and multimedia learning (Mason, Pluchino, & Tornatora, 2015, 2016; Mason, Scheiter, & Tornatora, 2017). EMME has been found to be beneficial for both the processing of information – using either texts, pictures, or medical images − and learning from it, that is, in terms of both processes and outcomes related to this type of modeling.

However, there are also a few studies documenting that EMME is not always more effective than traditional instruction when considering procedural knowledge for problem solving. For example, in the study by van Gog et al. (2009) students had to solve a procedural puzzle problem that was a mathematical problem. EMME showed how to solve it from the beginning to the end and was accompanied by verbal explanations. Results showed that EMME did not foster learning and in the transfer task EMME was even detrimental as students in the control condition outperformed those who watched the EMME video, likely because verbal explanations made EMME redundant in attention guiding (van Gog et al., 2009). In more recent research by van

Marlen, Vermerskerken, Jarodzka, & van Gog (2016), students were provided with EMME to solve a simple geometry. Findings revealed that EMME was effective only for the faster time taken to find solutions to transfer tasks by students who watched it (Experiment 1). When considering more complex geometry problems, EMME differentiated participants again only for the time spent in solving transfer problems, but in this case longer time was taken by those who observed EMME (Experiment 2).

In sum, despite a few studies on procedural problem-solving skills did not show favorable results after using EMME, the current literature most indicates that EMME results to be effective in promoting better processing of the examined information and related learning. One of the functions that explain the effectiveness of this type of video-modeling is attention guidance. EMME guides learners' attention to what the model is focusing on at that moment, and so the visual attention of the learner is guided by and synchronized with the model, as in a state of joint attention (van Marlen et al., 2018). This mechanism is particularly useful to account for the benefit of EMME when learners have to deal with transient information, like in a situation where a model explains information that is replaced once the model has finished talking about that information and continues to talk about another content, or when information is simply subtle and shortly available. In other words, students may lose relevant information when they do not pay attention to the right information at the right time (van Gog et al, 2009; van Marlen et al., 2018).

In the experimental study by Jarodzka et al. (2012), for example, two types of EMME were provided within an intervention to foster clinical reasoning in medical students through visual observations. EMMEs guided students' to pay attention to very subtle and short-term but relevant information regarding movement patterns in order to collect the right observations of symptoms that are crucial for a diagnosis of epileptic seizures in infants. In another experimental study by Jarodzka et al. (2013), EMMEs were used to guide attention in a visually complex perceptual task. Given difficulties in visual searching, realistic and dynamic videos were used to help learners focus on relevant fish locomotion patterns. Results showed that EMMEs were more effective than a traditional video without the model's eye movements in fostering learners' visual searching and interpretation of relevant information in classification tasks when presented with novel stimuli about fish locomotion.

Another function that explains the effectiveness of EMME is that it visualizes a perceptual and cognitive strategy that would otherwise remain unobservable for novice learners. For example, in Mason and colleagues' (2015, 2016, 2017) studies, students were shown an EMME of an expert who integrated text and picture while reading an illustrated text with where the information was much less transient. That is, EMME showed an entire visual strategy when learning from text and picture, which was picked up and learned by the observers. Findings revealed that students using EMME, compared to a control group, showed a stronger integrative processing of verbal and graphical information during the reading of a similar text, and higher learning outcomes after reading.

Interventions aimed at improving students' sourcing have mostly relied on worksheets, prompts, or group discussions to sustain the complex process that is part of critical reading as a systematic review has indicated (Brante & Strømsø, 2017). Based on the aforementioned literature on video modeling and video modeling through the gaze replay of an expert who performs well on a given task, it is theoretically legitimate to expect that EMME can be an effective way to improve sourcing as it shows an effective visual strategy to be implemented. The underlying reason is that seeing the model's strategy through various steps of source evaluation has a great potential to support a complex activity that is far from perceptual but involves perceptual processes (Mason et al., 2015, 2016, 2017). In fact, readers must identify and pay attention to source information and then spend time processing sources content according to their reliability. EMME provides a unique opportunity to visualize a perceptual strategy, which focuses on usually overlooked information such as “metadata” about source characteristics. As shown in eye-tracking studies, experts and novices differ in attention allocation; that is, compared to novices, experts pay attention to task-relevant information longer and faster, and they spend less time on task-irrelevant information (e.g., van Gog, Paas, & van Merriënboer, 2005). This difference in attention allocation may lead learners to fail to encode relevant source information while reading online digital documents (e.g., Brand-Gruwel et al., 2017).

As previous research has indicated, overlooking source information is common in students of different grade levels (e.g., Barzilai & Zohar, 2012; Clark & Slotta, 2000). To overcome this difficulty, EMME may represent a powerful tool as it shows the learner an effective strategy through the various steps of the sourcing process in a critical reading task. In the first steps, that is when inspecting a SERP snippets after a search and when reading the content of a webpage, a powerful strategy should be focused on identifying the source information that appears during our navigation on the Web. Compared to other modeling techniques, EMME has the unique affordance to model perceptual and cognitive strategies that are crucial for critical reading, but would otherwise remain unobservable for readers who have not acquired source evaluation skills.

The link between sourcing and comprehension of multiple documents on the same topic is not only theoretically justified, as mentioned above, but it is also empirically documented in studies revealing that attention to source information is associated with better multiple-document comprehension (e.g., Anmarkrud, Bråten, & Strømsø, 2014; Salmerón et al., 2010). Therefore, if EMME supports attention to source information, we can also expect it to sustain the comprehension of conflicting information provided by various sources. It is worth noting that in some studies EMME was also accompanied by verbal instructions, either unambiguous (e.g., van Gog et al., 2009) or ambiguous (van Marlen et al., 2019), to guide learners' attention, whereas in others, only EMME was provided, without any verbal supplement (Mason et al., 2015, 2016, 2017). Based on the latter, in the current study EMME was shown without any simultaneous verbal accompaniment in order to more clearly test its effectiveness in modeling a complex activity like sourcing. In this way, any benefit can only be attributed to the modeling itself. Moreover, participants are not asked to attend to both visual and verbal processes at the same time, which can be difficult. In this regard, unlike other studies that used verbal accompaniments, in the present study, the sourcing task represented in the EMME video did not involve visual search of subtle, transient elements that are available for a short time, which may not be perceived without verbal comments (Jarodzka et al., 2012).

Nevertheless, sourcing processes during Internet reading are complex because they involve paying attention to source information distributed across different sections (e.g., snippets in the SERP, author information on a webpage), they attend longer to information from a trustworthy source than from an untrustworthy source, and they use information from trustworthy sources to form an integrated representation of the topic. Thus, it is an open issue whether EMME, as an essentially perceptual tool, can also model a process that is far from perceptual but includes some perceptual aspects. However, findings on the effectiveness of EMME in modeling other non-perceptual processes, such as the integration of verbal and graphical information while reading an illustrated text without any verbal explanation, are encouraging (Mason et al., 2015, 2016, 2017).

The Study: Research Questions and Hypotheses

Sourcing is crucial when searching for information on the Internet and learning in online environments. The ability to identify and represent source information (metadata) to interpret a document's content and judge its authoritativeness, or to use source information in referring to a document's content, is essential in order to rule out or augment the content of messages on the basis of source credibility, that is, the trustworthiness and accuracy of the content (Goldman & Scardamalia, 2013; Salmerón et al., 2018). Given the importance of this ability in our information-saturated society, this study sought to extend current research on how to improve sourcing by adopting the EMME approach, which, to the best of our knowledge, has not been used in this area of research. The following research questions (RQ) guided the study:

(1) Would EMME increase the time spent reading (a) the SERP snippets, (b) webpage headers (i.e., the logo and the name of the institution hosting the webpage), (c) information about the text author within webpages (i.e., author's name and occupation), and (d) webpage texts?

(2) Would EMME increase (a) source citations and (b) number of ideas from web pages in post-test essays?

Based on the previously reviewed literature showing that EMME has positive effects on the processing of similar material by modeled learners (e.g., Jarodzka et al., 2013; Mason, 2015, 2016, 2017), for RQ1 we hypothesized that the EMME group would show more strategic processing of the online materials from pre to post-test, as reflected in attending longer to the snippets in the SERP (RQ1a), the webpage headers (RQ1b), and the authors' credentials (RQ1c). For RQ1, we also expected that students in the EMME group would selectively allocate their text processing times at post-test based on the webpage's trustworthiness, with higher reading times for trustworthy pages and lower reading times for untrustworthy pages (RQ1d).

For RQ2, we expected that the EMME group would cite more sources in their written essays from pre- to post-test, as a consequence of paying more attention to source information and tagging source features and contents (RQ2a) (Braasch & Bråten, 2017; Bråten et al., 2014; Stang Lund, Bråten, Brandmo, Brante, & Strømsø, 2014). We also hypothesized that the EMME group would report more ideas from web pages at post-test than at pre-test, particularly inter text inferences, as an effect of implicit evaluation elicited by the model, who strategically looked at logos and attended to source information (RQ2b). Students could use that "metadata" to reconcile the conflicting views given in different web pages (Stadtler & Bromme, 2014). The improvements in sourcing activities and products predicted for RQ 1 and 2 were not expected in the control group, which received no modeling.

Method

Participants

Sixty-four undergraduate students from a large Spanish university participated in the study (M_{age} = 20.8, SD = 1.91; 84.1% women). Most of the students were enrolled in their third (47.6%) or fourth (38.1%) year of undergraduate psychology or education programs (52.4 and 46.0%, respectively). Students volunteered either for class credit or for an economic compensation (10€). All participants signed an informed consent form and were debriefed after completing the study. From the original sample, we excluded 1 student due to incomplete data, which resulted in a final sample of 63 participants.

We performed a priori power analyses (G*Power 3; Faul, Erdfelder, Lang, & Buchner, 2007) to estimate the necessary sample size for our study. To our knowledge, there is no previous EMME intervention study like ours. Thus, we estimated the a priori effect size based on the results from three studies that performed other types of educational interventions aiming to improve undergraduate students' sourcing in a single session (Stadtler & Broome, 2007, 2008; Wiley et al, 2009). These three studies found several positive effects ranging from Hedges' $g = 0.66$ to 1.04 in several measures (for a review see Brante & Strømsø, 2018), such as the use of source information in essays ($g = 0.66$ to 1.04; Stadtler & Broom, 2007, 2008) or the time devoted to read reliable web pages ($g = 1.03$; Wiley et al., 2009). The average of these effect sizes can be considered as large (Cohen, 1988), so we performed the G*power analysis with an a priori effect size set at $f = .40$ (cut-off point for large effects from ANOVA) and with alpha and beta levels set at .05 and .20, respectively. The result indicated that a 52-participant sample was necessary to detect an interaction effect from a two-way mixed ANOVA. Therefore, our sample size ($N = 63$) can be considered as appropriate for our purposes.

Materials and Equipment

Webpages. Table 1 provides an overview of the main characteristics of the webpages used. Each participant read two separate sets of four webpages on socio-scientific conflicting topics. One set of webpages discussed pros and cons of the use of renewable energies as a potential solution to fight climate change (CC); the other discussed pros and cons of genetically modified food (GMF). The pages were assembled from various authentic online texts on the issue, including institutional and NGO reports, as well as diverse popular science articles. For each topic, two pages provided arguments in favor of and two against the main topic (use of renewable energies to fight climate change or genetically modified food). The level of trustworthiness of pages was manipulated by varying the degree of expertise and benevolence of the authors (Unkel & Hassel, 2017). Specifically, for pages providing a similar positive or negative view, one was authored by a trustworthy source (i.e. government agency or research institution) and another by an untrustworthy source (i.e. a company with commercial interests in the topic or laypersons writing personal blogs). Trustworthy webpages, but not untrustworthy ones, cited scientific sources in the text to support their main claims.

To ensure that the webpages were appropriate for undergraduate students, we computed readability scores for each webpage using the Flesch-Szigriszt Index (Szigriszt, 1992), which is a version of the classic Flesch Index in Spanish. The mean readability score for the CC webpages was 48.7 (*SD* = 6.1), and for the GMF webpages, 45.8 (*SD* = 0.8). According to the INFLESZ scale, values ranging between 40-55 correspond to the category “somewhat difficult”, which includes popular science texts or specialized press (Barrio-Cantalejo et al., 2008). This scale distinguishes five levels of text difficulty, ranging from “very difficult” (readability < 40; e.g., undergraduate textbooks) to "very easy" (readability > 80; e.g., primary school textbooks).

[Insert Table 1 about here]

EMME. We constructed five EMMEs that presented a dot to represent a student’s gaze on a SERP or on a particular page. Each EMME modeled a different strategy corresponding to advanced readers (see Table 2). Accordingly, participants were told that the EMMEs were from good students. Specifically, EMME 1 modeled a student who fully inspected the results from a SERP from top to the bottom by carefully reading each snippet; EMME 2 modeled a student who carefully inspected the source of information provided on a webpage, including the webpage header and information about the text author; EMME 3 modeled the strategy of deeply reading the text of a trustworthy page by slowing down his/her gaze movements and showing re-reading behavior, after attending to the source information; EMME 4 modeled the strategy of skimming the text of a less trustworthy page after attending to the source information; and EMME 5 modeled the strategy of quickly leaving a commercial page that was not topically relevant to the student's goal, after focusing on the source information.

[Insert Table 2 about here]

Essays. The average length of the essays was 235 words (*SD* = 65, *min* = 109, *max* = 420). Responses were analyzed in terms of sourcing and comprehension. We first divided each essay into ideas, defined as units with a main verb that expressed an event, activity, or state (Magliano, Trabasso, & Graesser, 1999). After segmentation, essays were coded to indicate whether the ideas contained an explicit reference to source information. Specifically, we identified if participants referred to the document source (i.e., author's occupation or the institution that hosted the page) or to embedded sources (i.e., scientific studies referenced in the webpages) (Salmerón, Gil, & Bråten, 2018b). Then, we identified the webpage or pages that contained that particular idea. Finally, ideas were coded to identify students' understanding of the topic. Specifically, we distinguished among three types of ideas: single idea paraphrases, intratext inferences, and intertext inferences. Single idea paraphrases included correct claims in which students expressed an idea from one of the pages in their own words (e.g., *If the Earth is meant to survive, it will be saved by itself, as always has been* [from I2C2 webpage] / *A researcher suggests that we let the Earth to renew itself* [from student essay]). Intratext inferences combined two single-idea paraphrases that were from one page, but not connected on the page (e.g., *Biofuel plantations, soybean in Latin America and oil palm in Indonesia, bring with them the deforestation of ancient forests. Deforestation produces an increase in carbon dioxide emissions to the atmosphere. This is because forests play a crucial role in climate change: they have the potential to absorb close to* *a tenth of the global carbon emissions projected for the first half of this century* [from personal blog]/ *Biodiesel is not good to combat global warming, as it needs enormous plantation spaces, producing enormous devastation of trees in mountains that are what help us to dissolve to a great extent the CO2 that humans emit* [from student essay]). Alternatively, in intratext inferences a paraphrase could be linked to some information from students' prior knowledge. Finally, intertext inferences combined two single-idea paraphrases from two different pages (e.g., *In my opinion, the use of biofuels instead of petroleum derivatives will not only not contribute to improving climate change* [from personal blog], and *The results confirm that using pure biodiesel, or mixed with conventional fuel, can reduce greenhouse gas emissions* [UNEP webpage] / *There is great controversy over whether biodiesel should be used, since according to some people it affects climate change, but others defend that it can be used to solve it* [from student essay]).

The first author and a trained research assistant, both blind to the conditions, independently scored a random selection of 12.7% of the essays for each of the topics: 7 essays included 101 idea units for the CC topic, and 7 essays contained 92 idea units for the GMF topic. The coding of students' references to sources yielded a Cohen's Kappa of 1 for both the CC and GMF topics, and an understanding of the content of .73 and .72 (for CC and GMF, respectively), thus showing substantial agreement. All disagreements were resolved through discussion between the two raters, and the research assistant scored the remaining essays according to the same coding systems.

Prior topic knowledge. Prior knowledge about the topics of climate change (CC) and genetically modified food (GMF) was assessed with a true-false measure shortened and adapted from prior studies (e.g., Salmerón et al., 2010). Both measures included items about scientific and political or historical issues for each topic. The internal consistency reliability for participants' scores was questionable (Cronbach's α = .60 and .62 for the CC (10 items) and GMF measure (9 items), respectively).

Topic interest. Participants' personal interest in the topics of CC and GMF was assessed by means of a questionnaire that asked them about their interest and active involvement in the issues, using a 10-point Likert-type scale ranging from 1 (not at all true of me) to 10 (very true of me). Both measures were shortened and adapted from prior studies (e.g., Bråten, Gil, Strømsø, & Vidal-Abarca, 2009). Cronbach's α for participants' answers on the CC measure (7 items) was excellent (α = .91), and it was good (α = .84) for answers on the GMF measure (6 items).

Equipment. We used a SMI REDn eye-tracker with a sample rate of 60 Hz. BeGaze software was used to extract fixations, using the low speed event detection algorithm with a minimum threshold for fixations set at 100 msec.

Procedure

Participants were tested individually in a computer lab. On arrival, they were randomly assigned to the experimental or control group. First, they completed a questionnaire on demographics and filled out the prior knowledge and interest measures. Then, they were introduced to the reading and writing task. They were told to imagine that they had a personal blog and wanted to write about the topic of CC/GMF. Specifically, they were told to read the pages to write a blog entry arguing about the pros and cons of possible solutions to fight climate change (CC topic) or genetically modified food (GMF topic). After being calibrated to the eye-tracking system (using a 9-point calibration), the Google SERP for the first topic was presented on the screen. From this SERP, they could access the four webpages. The participants were told that reading time was limited to seven minutes, but they were also free to end the task earlier. We based this time limit on pilot testing, which indicated that seven minutes would allow all the participants to read the four webpages and re-read some of them. Furthermore, a recent meta-analysis (Brysbaert, 2019) found that the average silent reading rate for adults in English is 238 words per minute (based on results from 144 studies), whereas in Spanish is even faster (278 words per minute, based on six studies). Thus, based on the reading results in Spanish, and given that the four texts used in our study were 1058-word long in total (ranging from 245 to 284), a total average time of 3.8 minutes was necessary to read them. A time limit of seven minutes was therefore long enough to read all the texts, also allowing rereading what the participants considered necessary[1]. In addition, one minute before the end of the set time the researcher warned the participants.

After they finished reading the webpages, participants were given 10 minutes to write the blog entry on a laptop using a word-processing application. They were also warned one minute before the end of the set time. Next, participants in the experimental group watched a series of EMMEs for 5 minutes. They were told that the videos displayed students' eyes while performing a multiple-document task similar to what they had been doing, and that the students were chosen because they were good performers. Participants in the control condition watched a 5-minute video about the topic they had just read and written about (either CC or GMF). Finally, all the students continued with the second topic for the post-test, undergoing the same procedure as on the pre-test. The session lasted approximately 90'.

Design

[1] In addition, the researchers that conducted the experimental sessions (first and second author) ascertained that all the participants read the texts within the given time-frame. Accordingly, eye-tracking results showed that participants took in average 5.15 and 4.96 minutes to read all the texts at pre-test and at post-test, respectively (maximum = 6.45 minutes at pre-test, and 6.61 at post-test).

We used a pre-post design with experimental and control groups. To minimize the impact of topic characteristics on the pre-post design, the presentation order was counterbalanced. In addition, to minimize the order effects from the results page, we created two SERPs with different rank orders for each topic.

Results

Preliminary analyses

As a first step in data preprocessing, raw fixation duration time data were examined to detect outliers, i.e. individual fixations that lasted 2 SD above or below each student's fixation duration mean. On average, outlier fixations represented 4.28% of individual fixations. They were replaced by the student's median of fixation duration times. Next, fixations were aligned with the corresponding area of interest (AOI), which included: (a) individual snippets from the SERP page, (b) the webpage hosting the institution's logo and name, corresponding to the header of each page, (c) webpage text corresponding to the paragraphs on each page, and (d) the text authors' names and affiliations included on each page in a location clearly separated from the text. Fixations times for the AOI with textual information (i.e. SERP's snippets, text authors' names and affiliations, and paragraphs of each page) were converted to time-per-character measures (ms/char) to control for the differences in text length. Fixations times for the non-textual AOI (i.e. logo of each page) were not transformed.

We also checked whether the two groups differed in their prior knowledge and interest in the two topics used in the study. An ANOVA with topic (CC and GMF) and condition (EMME vs. control) and the percentage of correct responses on the prior knowledge tests revealed non-significant effects of topic, condition, or the interaction (all Fs < 1). A similar analysis with topic interest as dependent variable revealed a large significant effect of topic, $F(1, 61) = 20.91$, $p < .001$, $\eta^2_p = .26$, and non-significant effects of condition (both *F*s < 1). Participants showed slightly higher interest in CC (*M* = 7.3, *SD* = 1.2) than in GMF (*M* = 7.1, *SD* = 1.3). It is important that we counterbalanced the topics across time and condition to control for the potential effect of topic interest. More critically, prior knowledge and interest in the topic did not differ between conditions.

Processing of Source and Textual Information

Distributions of eye-movements indices at pre- and post-test were inspected for normality. Skewness and kurtosis values were in the range of -2 and +2 (see Table 3), indicating that the data followed a normal distribution (Field, 2009; George & Mallery, 2010).

To test the various hypotheses related to RQ1, we examined students' online processing of source information, including (a) SERP snippets, (b) webpage headers, and (c) information about the text authors. First, we performed a mixed ANOVA with condition (control and EMME) as between-participant variable, time (pre and post) as within-participant variable, and average reading time of the SERP snippets as dependent variable. Results revealed significant effects of condition, $F(1, 61) = 5.67$, $p = .02$, $\eta^2_p = .09$, and time, $F(1, 61) = 7.26$, $p < .01$, $\eta^2_p = .11$. These results were qualified by a significant interaction between condition and time, $F(1, 61) = 14.75$, $p < .001$, $\eta^2_p = .20$. Post-hoc contrasts with Bonferroni correction indicated differences between conditions at post-test ($p < .001$), but not at pre-test ($p = .88$). Participants in the EMME group read the SERP snippets longer than those in the control group only at post-test (see Table 3). Across time, participants in the control group did not differ ($p = .42$), whereas those in the EMME group increased their SERP reading times from pre to post-test ($p < .001$).

We further explored students' processing of source information within the webpages. First, we performed a mixed ANOVA with condition (control and EMME) as between-participant variable, time (pre- and post-test) as within-participant variable, and average time spent on the webpage header as dependent variable. Results revealed a significant effect of time, $F(1, 61) = 8.04$, $p < .01$, $\eta^2_p = .12$, and a non-significant effect of condition or the interaction (both *Fs*<1). Participants in both groups inspected the headers for a longer time at post-test than at pre-test.

Second, we performed a mixed ANOVA with condition (control and EMME) as between-participant variable, time (pre and post-test) as within-participant variable, and average time spent on the webpage's information about the text author as dependent variable. Results revealed non-significant effects of time, $F(1, 61) = 2.08$, $p = .16$, $\eta^2_p = .03$, and condition (F<1), and a significant interaction, $F(1, 61) = 6.08$, $p = .02$, $\eta^2_p = .09$. Post-hoc analyses with Bonferroni correction indicated that, across time, the control group did not differ in the time spent attending to text author information ($p = .48$), whereas the EMME group increased their times from pre- to post-test ($p < .01$). Across groups, there were no differences at pre-test ($p = .13$) or at post-test ($p = .22$). Whereas participants in the EMME group, compared to those in the control group, spent more time attending to text author information at post-test, this difference failed to reach significance levels.

Finally, for the analysis of text reading times (RQ1d), we included page trustworthiness as an additional factor because in our hypothesis we predicted that the experimental effects would be conditional on this factor related to source evaluation. Thus, we performed a mixed ANOVA with condition (control and EMME) as between-participant variable, time (pre and post) and page trustworthiness (high or low) as within-participant variables, and average text reading time as dependent variable. Results revealed significant effects of time, $F(1, 61) = 8.38$, $p < .01$, $\eta^2_p = .12$, and page trustworthiness, $F(1, 61) = 19.00$, $p < .001$, $\eta^2_p = .24$, but not condition, F<1. There were also significant two-way interactions between time and page trustworthiness, $F(1, 61) = 4.14$, $p = .04$, $\eta^2_p = .06$, and between page trustworthiness and condition, $F(1, 61) = 4.74$, $p = .03$, $\eta^2_p = .07$, but not between time and condition, $F(1, 61) = 1.97$, $p = .17$, $\eta_p^2 = .03$. These effects were qualified by a two-way interaction between condition and page trustworthiness, $F(1, 61) = 4.74$, $p = .03$, $\eta^2_p = .07$, a two-way interaction between time and page trustworthiness, $F(1, 61) = 4.14$, $p = .04$, $\eta^2_p = .06$, and a three-way interaction between time, page trustworthiness, and condition, $F(1, 61) = 6.62$, $p = .01$, $\eta^2_p = .10$.

We then conducted post-hoc, Bonferroni-corrected comparisons to further examine the effects above. These analyses indicated that whereas the intervention groups showed similar text reading times at pre-test regardless of the webpages trustworthiness (trustworthy: $p = .78$; untrustworthy: $p = .76$), there were differences at post-test. At this time, the EMME group devoted lesser time than the control group reading the untrustworthy pages, $p = .03$, and more time reading the trustworthy pages, although in this case the difference did not reach significance, $p = .06$. Accordingly, the results showed that the EMME group reduced the reading time of the texts from untrustworthy pages at post-test, as compared to the pre-test, $p < .001$, which was not the case for the control group, $p = .85$. With respect to the trustworthy pages, although the EMME group increased the text reading times at the post-test and the control group decreased it, as compared to the pre-test, the differences were not significant in any case, $p = .19$ and $p = .40$. Finally, these results were driven by the fact that, at pre-test, both groups read the trustworthy pages faster that the untrustworthy pages, both *p*s = .001, as was also the case for the control group at post-test, $p = .001$. However, at post-test the EMME group read both types of pages for a similar time, $p = .36$, which was even slightly longer for the trustworthy pages (see Table 4).

In sum, the results provided partial confirmation of our hypotheses related to RQ1a and RQ1c. From pre- to post-test, students in the EMME group, but not those in the control group, increased their processing time of SERP snippets and text author information within the webpages. No differences were observed in attention to webpage headers (RQ1b). Results for RQ1d, although less clear, partially support our hypothesis. Whereas participants in the control group did not change their text reading times based on the webpage’s trustworthiness from pre- to post-test, those in the EMME group shifted from reading trustworthy pages faster than untrustworthy pages at pre-test to reading both types of pages at a similar rate at post-test. More critically, at post-test, participants in the EMME group showed the modeled strategies in reading texts depending on their level of trustworthiness. That is, compared to those in the control group, they read untrustworthy pages at a quicker rate.

Source Citations and Quality of Ideas in Written Essays

We inspected distributions of essay indices at pre- and post-test. Skewness and Kurtosis values of intra and inter-text inference ideas were in the range of -2 and +2, and therefore the distribution was considered appropriate for parametric analyses (see Table 3). Distribution of single paraphrase ideas indicated non-normal values. A logarithmic transformation of the data was needed to fit the scores into a normal distribution (Skewness at pre-test = .88, at post-test = .89; Kurtosis at pre-test = -.13, at post-test = -.11). Finally, scores for citations of document and embedded sources indicated large deviations from normality. Close inspection of the data revealed that a majority of participants did not include any citation in their essays. Across groups, the percentage of students whom did not cite at all ranged from 56.25 to 74.19% for document sources, and from 48.39 to 67.74% for embedded sources. Accordingly, we decided to transform scores for each type of source in Boolean data, indicating if participants cited (1) or not (0) at least one source for a particular category.

To test the effects of EMME on source citations (RQ2a), we ran non-parametric tests with condition (control or EMME) as between-participant variable, and time (pre and post) as within-participant variable, for the scores citation of document and embedded sources (yes or no). To test the between-effects we used Mann-Whitney tests comparing condition. Results indicated that at pre-test participants did not differ for the presence of citation of document (U = 476.5, p = .72, d = .17) or embedded sources (U = 408.0, p = .16, d = .40). At post-test, the number of participants citing document sources was higher in the EMME than in the control groups, although such difference was not statistically significant (U = 391.00, p = .08, d = .47), as was the case for embedded sources (U = 472.57, p = .71, d = .18). Next, to test the within-effects we used the McNemar test, comparing pre to post-test scores in each group. In the control group, participants did not change their citation patterns from pre to post-test, neither for document (p = 1) nor for embedded sources (p = .21). In the EMME group, a higher number of participants cited document sources at post than at pre-test (p < .04), while no difference was observed for the citation of embedded sources (p = 1). Specifically, in the EMME group, 8 students out of the 25 who had not cited a document source at pre-test cited at least one document source at post-test.

Next, to test potential effects of EMME on the quality of the ideas included in the essays (RQ2b), we computed a mixed ANOVA with condition (control or EMME) as between-participant variable, and time (pre and post) and idea type (single paraphrase [log transformed], intra, or inter text inferences) as within-participant variables. Because we introduced a new scale for the single paraphrase scores (i.e. log transformation), we also standardized the scores of the three types of ideas to make them comparable. There were non-significant main effects of condition, time, or idea type (all three *Fs* < 1). Similarly, we found non-significant interactions between condition and time, *F*<1, condition and idea type, $F(2, 61) = 2.16$, $p = .12$, $\eta^2_p = .03$, time and idea type, *F*<1, or the three-way interaction, *F*<1.

In summary, the results provide only limited support for our hypotheses on the impact of EMME on students' essays. Regarding the effects of EMME on source citations, our predictions were partially confirmed. Specifically, EMME, and not the control group, significantly increased the number of participants who moved from not citing document sources at pre-test, to cite at least one document source at post-test. However, the difference between the percentages of participants who cited document sources at post-test, although in the expected direction, was not statistically detectable when comparing the control and EMME groups. Finally, our predictions regarding the effect of EMME on post-test essays ideas was not confirmed, as EMME had no effect on the number of ideas (single paraphrase, intra, or inter text inferences) included in the students' post-test essays.

Discussion

This study identified, for the first time, the positive effects of a short instruction based on EMME on students' critical evaluation of information sources when reading multiple webpages about a conflicting topic. Moreover, the use of eye-tracking methods allowed us to determine how EMME affected students' processing (i.e., attention to source information) and outcomes (i.e., essay quality).

By watching a series of EMME that modeled attention to source information on SERPs and webpages, undergraduate students increased their visual attention to SERP snippets and text author information within the webpages, but not to the webpages' headers. Changes in the way the SERP was processed were particularly noteworthy (medium to high effect size), indicating that participants in the EMME group showed an increased "click restraint", moving from a novice to an expert strategy (Brand-Gruwel et al., 2017; Wineburg & McGrew, 2017). Because SERPs synthesize critical source information from all the webpages in a single space, participants could have used them to reflect on the relevance and quality of the pages and directly compare them (Kammerer & Gerjets, 2014; Salmerón et al., 2010). Corroborating this assumption, our results showed that EMME increased the probability to cite document sources from pre to post-test essays'.

The effects of EMME on students' attention to text author information within the webpages are less straightforward. EMME increased students' attention to author information, but it had no impact on the processing of webpage headers. In this learning context, source information was distributed across several pages (i.e. SERPs and individual webpages), and source information provided by SERPs was partially redundant with information from the webpages, particularly the information on the headers. The high increase in SERP processing times in the EMME group could have kept students from processing further information that was redundant.

EMME also changed the way students read the texts. Specifically, participants in the EMME group increased their reading times of texts from trustworthy webpages, but not from untrustworthy webpages. The control group did not change their text reading times. In addition, at post-test, the EMME group tended to devote more time to reading trustworthy pages and significantly spent less time to untrustworthy pages than the control group. These effects are important, given the inherent difficulty of conveying a complex sequence of strategic decisions in EMME. Three videos from our study modeled selective reading in five steps, but only the first, second, and fifth steps were salient (carefully reading the SERP, looking at source information, and skipping a nonrelevant webpage, respectively) because the third (deeply reading a trustworthy text) and fourth (skimming an untrustworthy text) were implicit. Specifically, in order to perform these implicit steps, students first had to observe that the model attended to source information. Second, they had to recognize whether the page was trustworthy or not. Third, they had to note that the model modified their reading pace (slow reading and rereading of trustworthy tests, and quick skimming of untrustworthy texts). Fourth, they had to infer that the model adjusted the reading pace to the level of trustworthiness of the page.

In sum, participants were able to infer the strategic processing conveyed in the complex EMME. Importantly, participants benefited from this strategic behavior, as EMMEs increased the probability to cite document sources from pre to post-test essays’, such as author’s occupation or the institution that hosted and published the page. In the context of multiple document comprehension, sourcing in essays is an essential process as it reflects readers’ efforts to organize and discuss the different perspectives of a particular topic (Braasch & Bråten, 2017; Britt & Aglinskas, 2002).

Nevertheless, EMME failed to improve participants’ comprehension, as indicated by the null effects on the ideas included in their essays. This pattern of results opens up two different avenues for future research. First, research could analyze how students’ individual differences, for example, in reading comprehension, interact with an instruction based on EMME, because more advanced students could profit from them to a greater extent. Second, different EMME designs can be explored, for example, by adding verbal self-explanations to model the steps that cannot be conveyed visually (Salmerón & Llorens, 2019).

Together, our results support the need to go beyond outcome measures in educational research and identify process measures as essential components in order to fully understand complex educational scenarios, such as reading multiple webpages on the Internet. As Harteis, Kok, and Jarodzka (2018) have pointed out, including process measures in education studies allow researchers to understand how students learn, and not just what they learn.

Educational Implications

As pointed out in previous investigations, EMME is not only a research tool but also an instructional tool because it can be used in classrooms to model the use of effective strategies for successful performance on complex tasks and activities (Mason et al., 2016). To produce a video with the model's eye movements is costly to some extent, as it requires a device that register visual behavior. Although eye trackers are becoming less and less expensive, they are not cheap. However, once the videos are prepared, they can easily be used by teachers and instructors when introducing a new strategy or sequence of strategies. Unlike any other means, videos as those used in our study model a perceptual strategy that is made observable and students can pick up. Specifically, they can learn an entire visual sourcing strategy, which helps them to perform better in another searching context. Videos with an instructor who explains, or points to what to pay attention to, does not model a visual strategy with its unique affordances to sustain a complex activity such us sourcing.

Teachers and instructors can highlight that, even for complex reading tasks that require far more than perceptual processes, the latter are at the basis of information processing. In other words, EMME can contribute to refining students' metacognitive awareness of the various steps into which a complex process can be divided, and the specific strategies that are powerful in completing a task or activity in a learning context.

Paying attention to relevant elements through perceptual cues – like gaze replays – for appropriate encoding of information is the first step in moving toward successful performance on complex and demanding tasks (Jarodtzka et al, 2013; van Marlen et al., 2018). Source evaluation for critical reading is one of these tasks that require visual attention to particular information in order to discriminate between trustworthy and untrustworthy sources. Even short videos showing the visual behavior of a successful performer seem to be effective, at least to some extent, for both the processes and outcomes when reading conflicting documents on the same topic. Therefore, EMME can be implemented in a relatively short time.

Moreover, EMMEs can represent a starting point of tutorials, particularly in blended and online learning environments, and they can also take advantage of learning at one's own time and pace (e.g., Rienties, Tempelaar, Nguyen, & Littlejohn, 2019). Through attentional guidance, EMME contributes to making some essential aspects salient to students in order to increase their online and offline performance, as revealed by process and outcome measures. Even tasks that involve higher-order thinking processes, such as critical reading, are based on appropriate encoding of relevant information during effective processing of text content.

Limitations

This study has some limitations. The first is that the design included only two conditions. Eye-tracking studies with complex learning materials are laborious, and practical constraints do not always allow optimal research designs. However, a stronger investigation on the potential of EMME to enhance source evaluations should add more control conditions. Future investigations can benefit, for example, from the inclusion of a condition characterized by a more traditional video-modeling that offers visual cueing pointing to relevant source information by means of arrows. In an alternative control condition, EMME can be combined with the generation of observers' self-explanations.

The second limitation is the use of only four webpages (one text each) for each topic. In the future, more complex designs should also include a set of documents from sources that vary in trustworthiness to a more subtle degree, so that it becomes more challenging to differentiate among the sources. Related to this, the third limitation is that we assumed the trustworthiness of the four webpages on the basis of the sources' characteristics, but we do not know whether the participants perceived the sources' characteristics in the same way. However, we are inclined to assume that university students are able to distinguish between a webpage of the World Health Organization (WHO) and a webpage of a company like Monsanto, or a page of the United Nations Environment Program (UNEP) and a personal blog, when they encode source information. Nevertheless, future investigations will benefit from a manipulation check regarding webpages' trustworthiness.

The fourth limitation is that we considered only textual information. However, when we search the Internet for information, we face multimedia materials that introduce static or dynamic visualizations. It seems worthwhile to investigate attentional guidance in relation to both types of representations, as well as the integration of verbal and graphic information on pages with different levels of trustworthiness.

Conclusions

Despite these limitations, this study is the first to indicate that a short EMME can be used to foster important aspects of source evaluation processes for critical reading of documents on debated issues. Attention to the results that appear on a SERP after a search, and to source information within the accessed pages, can be increased through EMME. Moreover, the study also provides evidence that EMME play a role in post-reading essays about the contents read online − in terms of the document sources cited. Thus, EMME are to some extent effective in enhancing, either directly or indirectly, the processing of the online reading material, which then has beneficial consequences in terms of the sources discussed in their essays when engaging in critical reading.

References

Anmarkrud, O., 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. doi:10.1016/j.lindif.2013.01.007

Bandura, A. (1977). *Social learning theory*. Englewood Cliffs, NJ: Prentice-Hall.

Barrio-Cantalejo, I. M., Simón-Lorda, P., Melguizo, M., Escalona, I., Marijuán, M. I., & Hernando, P. (2008). Validación de la Escala INFLESZ para evaluar la legibilidad de los textos dirigidos a pacientes [Validation of the INFLESZ scale to evaluate the readability of patient texts]. *Anales del Sistema Sanitario de Navarra, 31,* 135-152.

Barzilai, S., Tzadok, E., & Eshet-Alkalai, Y. (2015). Sourcing while reading divergent expert accounts: Patways from views of knowing to written argumentation. *Instructional Science, 43*, 737-766. doi:10.1007/s11251-015-9359-4

Barzilai, S., & Zohar, A. (2012). Epistemic thinking in action: Evaluating and integrating online sources. *Cognition and Instruction, 30,* 39-85. doi:10.1016/j.learninstruc.2010.02.008

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. doi:10.1080/00461520.2017.1323219

Braasch, J. L. G., Bråten, I., Strømsø, H. I., Anmarkrud, O., & Ferguson, L. E. (2013). Promoting secondary school students' evaluation of source features of multiple documents. *Contemporary Educational Psychology, 38*, 180-195. doi:10.1016/j.cedpsych.2013.03.003

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. doi:10.1111/jcal.12162

Brand-Gruwel, S., & Wopereis, I. (2006). Integration of the information problem-solving skill in an educational programme: The effects of learning with authentic tasks. *Technology, Instruction, Cognition, and Learning, 4,* 243-263.

Brante, E. W., & Strømsø, H. I. (2018). Sourcing in text comprehension: A review of interventions targeting sourcing skills. *Educational Psychology Review, 30*, 773-799. doi:10.1007/s10648-017-9421-7

Bråten, I., Braasch, J., & Salmerón, L. (2019). Reading multiple and non-traditional texts: new opportunities and new challenges. In E. B. Moje, P. Afflerbach, P. Enciso, & N. K. Lesaux (Eds.), *Handbook of Reading Research* (Vol. 5). New York: Routledge.

Bråten, I., Ferguson, L. E., Strømsø, H. I., & Anmarkrud, Ø (2014). Students working with multiple conflicting documents on a scientific issue: Relations between epistemic cognition while reading and sourcing and argumentation in essays. *British Journal of Educational Psychology, 84*, 58-85. doi:10.1111/bjep.12005

Bråten, I., Gil, L., Strømsø, H. I., & Vidal-Abarca, E. (2009). Personal epistemology across cultures: Exploring Norwegian and Spanish university students' epistemic beliefs about climate change. *Social Psychology of Education, 12*, 529-560. doi:10.1007/s11218-009-9097-z

Bråten, I., Strømsø, H. I., & Andreassen, R. (2016). Sourcing in professional education: Do text factors make any difference? *Reading and Writing, 29,* 1599-1628. doi: 10.1007/s11145-015-9611-y

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

Brysbaert, M. (2019). How many words do we read per minute? A review and meta-analysis of reading rate. *Journal of Memory and Language, 109*, 104047. doi:10.1016/j.jml.2019.104047

Clark, D. B., & Slotta, J. D. (2000). Evaluating media-enhancement and source authority on the Internet: the Knowledge Integration Environment. *International Journal of Science Education, 22*(8), 859-871. doi:10.1080/0950069004123

Delgado, P., Avila, V., Fajardo, I., & Salmerón, L. (2019). Training young adults with intellectual disability to read critically on the Internet. *Journal of Applied Research in Intellectual Disabilities, 32,* 666-677. doi:10.1111/jar.12562

Faul, F., Erdfelder, E., Lang, A., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods, 39,* 175-191. doi:10.3758/BF03193146

Frerejean, J., van Strien, J. L. H., Kirschner, P. A., & Brand-Gruwel, S. (2018). Effects of a modelling example for teaching information problem solving skills. *Journal of Computer Assisted Learning, 34*, 688-700. doi:10.1111/jcal.12276

Groenendijk, T., Janssen, T., Rijlaarsdam, G., & van den Bergh, H. (2013). The effect of observational learning on students' performance, processes, and motivation in two creative domains. *British Journal of Educational Psychology, 83*, 3-28. doi:10.1111/j.2044–8279.2011.02052.x

Harteis, C., Kok, E. M., & Jarodzka, H. (2018). The journey to proficiency: Exploring new objective methodologies to capture the process of learning and professional development. *Frontline Learning Research, 6*, 1-5. doi:10.14786/flr.v6i3.435

Hefter, M. H., ten Hagen, I. Krense, C., Berthold, K., & Renkl, A. (2019). Effective and efficient acquisition of argumentation knowledge by self-explaining examples: videos, texts, or graphic novels? *Journal of Educational Psychology*. Advance online publication. doi:10.1037/edu0000350

Hoogerheide, V., Renkl, A., Fiorella, L., Paas, & van Gog, T. (2019). Enhancing example-based learning: Teaching on video increases arousal and improves problem-solving performance. *Journal of Educational Psychology, 111*, 45-56. doi.org/10.1037/edu0000272

Jarodzka, H., Balslev, T., Holmqvist, K., Nyström, M., Scheiter, K., Gerjets, P., & Eika, B. (2012). Conveying clinical reasoning based on visual observation via eye-movement modelling examples. *Instructional Science*, *40,* 813-827. doi:10.1007/s11251-012-9218-5

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. doi:10.1111/jcal.12189

Jarodzka, H., van Gog, T., Dorr, M., Scheiter, K., & Gerjets, P. (2013). Learning to see: Guiding students' attention via a Model's eye movements fosters learning. *Learning and Instruction*, *25,* 62-70. doi:10.1016/j.learninstruc.2012.11.004

Kammerer, Y., Bråten, I., Gerjets, P., & Strømsø, H. I. (2013). The role of Internet-specific epistemic beliefs in laypersons source evaluations and decisions during Web search on a medical source. *Computers in Human Behavior, 29*, 1193-1203. doi:10.1016/j.chb.2012.10.012

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. doi:10.1080/10447318.2013.846790

Kammerer, Y, Meier, N., & Stahl, E. (2016). Fostering secondary-school students' intertext model formation when reading a set of websites: The effectiveness of source prompts. *Computers & Education*, *102,* 52-64. doi:10.1016/j.compedu.2016.07.001

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

Kuiper, E., Wolman, M., & Terwel, J. (2008). Integrating critical Web skills and content knowledge: Development and evaluation of a 5th grade educational program. *Computers in Human Behavior*, *24*, 666-692. doi:10.1016/j.chb.2007.01.022

List, A., Du, H., Wang, Y., & Lee, H. Y. (2019). Toward a typology of integration: Examining the documents model framework. *Contemporary Educational Psychology, 58,* 228-242. doi:10.1016/j.cedpsych.2019.03.003

Litchfield, D., Ball, L. J., Donovan, T., Manning, D. J., & Crawford, T. (2010). Viewing another person's eye movements improves identification of pulmonary nodules in chest X-ray inspection. *Journal of Experimental Psychology: Applied, 16,* 251-262. doi:10.1037/a0020082

Macedo-Rouet, M., Braasch, J. L. G., Britt, A. M., & Rouet, J-F. (2013). Teaching fourth and fifth graders to evaluate information sources during text comprehension. *Cognition and Instruction, 31*, 204-226. doi:10.1080/07370008.2013.769995

Macedo-Rouet, M., Potocki, A., Scharrer, L., Ros, C., Stadtler, M., Salmerón, & Rouet, J-F. (2019). *How good is this page? Benefits and limits of prompting on adolescents' evaluation of web information quality*. Advance online publication. doi:10.1002/rrq.241

Magliano, J. P., Trabasso, T., & Graesser, A. C. (1999). Strategic processes during comprehension. *Journal of Educational Psychology, 91*, 615-629. doi:10.1037/0022-0663.91.4.615

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

Mason, L., Pluchino, P., & Tornatora, M. C. (2015). Eye-movement modeling of text and picture integration during reading: effects on processing and learning. *Contemporary Educational Psychology, 41*, 172-187. doi:10.1016/j.cedpsych.2015.01.004

Mason, L., Pluchino, P., & Tornatora, M. C. (2016). Using eye-tracking technology as an indirect instruction tool to improve text and picture processing and learning. *British Journal of Educational Technology, 47*, 1083-1095. doi:10.1111/bjet.12271

Mason, L., Scheiter, K, & Tornatora, M. C. (2017). Using eye movements to model the sequence of text-picture processing for multimedia comprehension *Journal of Computer Assisted Learning, 33,* 443-460. doi:10.1111/jcal.12191

Paul, J., Cerdán, R., Rouet, J. F., & Stadtler, M. (2018). Exploring fourth graders' sourcing skills/Un análisis de la capacidad de escrutinio sobre las fuentes de información de los estudiantes de cuarto grado. *Infancia y Aprendizaje, 41,* 536-580. doi:10.1080/02103702.2018.1480458

Paul, J., Stadtler, M., & Bromme, R. (2019). Effects of a sourcing prompt and conflicts in reading materials on elementary students' use of source information. *Discourse Processes, 56,* 155-169. doi:10.1080/0163853X.2017.1402165

Preacher, K.J., & Hayes, A.F. (2008). Asymptotic and resampling strategies for assessing and comparing indirect effects in multiple mediator models. *Behavior Research Methods, 40,* 879-891. doi:10.3758/BRM.40.3.879

Renkl, A. (2014). Toward an instructionally oriented theory of example-based learning. *Cognitive Science, 38*, 1-37. doi:10.1111/cogs.12086

Rienties, B., Tempelaar, D., Nguyen, Q., & Littlejohn, A. (2019). Unpacking the intertemporal impact of self-regulation in a blended mathematics environment. *Computers in Human Behavior*. Advance online publication. doi:10.1016/j.chb.2019.07.007

Salmerón, L., Gil, L., & Bråten, I. (2018a). Using eye-tracking to assess sourcing during multiple document reading: A critical analysis. *Frontline Learning Research, 6,* 105-122. doi:10.14786/flr.v6i3.368

Salmerón, L., Gil, L., & Bråten, I. (2018b). Effects of reading real versus print-out versions of multiple documents on students' sourcing and integrated understanding. *Contemporary Educational Psychology, 52,* 25-35. doi:10.1016/j.cedpsych.2017.12.002

Salmerón, L., Gil, L., Bråten, I., & Strømsø, H. (2010). Comprehension effects of signalling relationships between documents in search engines. *Computers in Human Behavior*, *26*, 419-426.

Salmerón, L., & Llorens, A. (2019). Instruction of digital reading strategies based on eye-movements modeling examples. *Journal of Educational Computing Research, 57,* 343-359. doi:10.1177%2F0735633117751605

Salmerón, L., Macedo-Rouet, M., & Rouet, J.-F. (2016). Multiple viewpoints increase students' attention to source features in social question and answer forum messages. *Journal of the Association for Information Science and Technology, 67,* 2404-2419. doi:10.1002/asi.23585

Salmerón, L., Strømsø, H. I., Kammerer, Y., Stadtler, M., & van den Broek, P. (2019). Comprehension processes in digital reading. In M. Barzillai, J. Thomson, S. Schroeder, & P. van den Broek (Eds.). *Learning to read in a digital world (pp. 91-120).* Amsterdam, John Benjamins.

Seppänen, M., & Gegenfurtner, A. (2012). Seeing through a teacher's eye improves students' imagining interpretation. *Medical Education*, *46*, 1113-1114. doi:10.1111/medu.12041

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

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

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. Braasch (Eds.), *Processing inaccurate information: Theoretical and applied perspectives from cognitive science and the educational sciences* (pp. 379–402). Cambridge, MA: MIT Press.

Stadtler, M., Paul, J., Globoschütz, S., & Bromme, R. (2015). Watch out! - An instruction raising students' epistemic vigilance augments their sourcing activities. In: D. C. Noelle, R. Dale, A. S. Warlaumont, J. Yoshimi, T. Matlock, C. D. Jennings, & P. P. Maglio, (Eds.), *Proceedings of the 37th Annual Conference of the Cognitive Science Society* (pp. 2278-2283). Austin, TX: Cognitive Science Society.

Strømsø, H. I., Bräen, I., & Britt, A. M. (2010). Reading multiple texts about climate change: The relationship between memory for sources and text comprehension. *Learning and Instruction, 20,* 192-204. doi:10.1016/j.learninstruc.2009.02.001

Szigriszt, F. (1992). *Sistemas predictivos de legilibilidad del mensaje escrito: Fórmula de perspicuidad.* Unpublished doctoral dissertation, Universidad Complutense de Madrid.

Tarchi, C., & Mason, L. (2019). Effects of critical thinking on multiple-document comprehension. *European Journal of Psychology of Education*. Advance online publication. doi:10.1007/s10212-019-00426-8 van Gog, T., Jarodzka, H., Scheiter, K., Gerjets, P., & Paas, F. (2009). Attention guidance during example study via the model's eye movements. *Computers in Human Behavior, 25*, 785–791. doi:10.1016/j.chb.2009.02.007

van Gog, T., Paas, F., & Van Merriënboer, J. J. G. (2005). Uncovering expertise-relateddifferences in troubleshooting performance: Combining eye movement and con-current verbal protocol data**.** *Applied Cognitive Psychology, 19***,** 205–221.https://doi.org/10.1002/acp.1112

van Marlen, T., Wermeskerken, M., Jarodzka, H., & van Gog, T. (2018). Effectiveness of eye movement modeling examples in problem solving: The role of verbal ambiguity and prior knowledge. *Learning and Instruction, 58,* 274-283. doi:10.1016/j.learninstruc.2018.07.005

van Marlen, T., Van Wermeskerken, M., Jarodzka, H., & Van Gog, T. (2016). Showing a model's eye movements in examples does not improve learning of problem-solving tasks. *Computers in Human Behavior*, *65*, 448–459. doi:10.1016/j.chb.2016.08.041

VanSledright, B., & Kelly, C. (1998). Reading American history: The influence on multiple sources on six fifth graders. *The Elementary School Journal, 98,* 239-265.

Wiley, J., Goldman, S., Graesser, A., Sanchez, C., Ash, I. & Hemmerich, J. (2009). Source evaluation, comprehension, and learning in internet science inquiry tasks. *American Educational Research Journal, 46,* 1060-1106. doi:10.3102%2F0002831209333183

Wineburg, S. (1991). Historical problem solving: A study of the cognitive processes used in the evaluation of documentary and pictorial evidence. *Journal of Educational Psychology*, *83,* 73-87. doi:10.1037/0022-0663.83.1.73

Wineburg, S. & McGrew, S. (2017). Lateral reading: Reading less and learning more when evaluating digital information. *Stanford History Education Group Working Paper No. 2017-A1*. doi:10.2139/ssrn.3048994

Table 1

Overview of the webpages for the two topics.

Topic	Webpage	Author	Content	Number of words
GMF	World Health Organization (WHO)	Director of Environmental Health of WHO	Argues in favor of GMF. States that GMF contain more vitamins and minerals and can be better conserved. The text cites a scientific study to support this view.	272
	Monsanto (Global Agriculture Company)	Director of department of risk management from Monsanto	Argues in favor of GMF. Outlines higher resistance to insects or diseases of GM crops and higher productivity.	271
	Food and Agriculture Organization of the United Nations (FAO)	Spanish representative of FAO	Argues against GMF. Discusses ecological risks of GM crops for surrounding soil and plants. The text cites a scientific study to support this view.	265
	Personal blog	Economist	Argues against GMF. State potential toxicity of GM proteins.	279
CC	Iberdrola (Spanish multinational electric utility company)	Iberdrola’s project director	The company sector addressed with renewable energy business. Argues in favor of solutions to CC. States the advantages of wind, solar, and hydroelectric power.	252
	United Nations Environment Program (UNEP)	Director of UNEP	Argues in favor of using renewable energies as a solution to CC. Discusses reduction of carbon dioxide emissions by using biodiesel. The text cites a scientific study to support this view.	271
	Personal blog	Undergraduate student in History	Argues against using renewable energies. Outlines that biofuels make CC worse.	269
	I2C2 (Spanish climate change research institute)	Communication director of Research Institute	Argues against using renewable energies. Discusses great economic costs and little success of renewable energies. The text cites a scientist from Harvard university to support the claims raised.	296

Note. GMF = Genetically Modified Food; CC = Climate Change

Table 2

Overview of the EMMEs used.

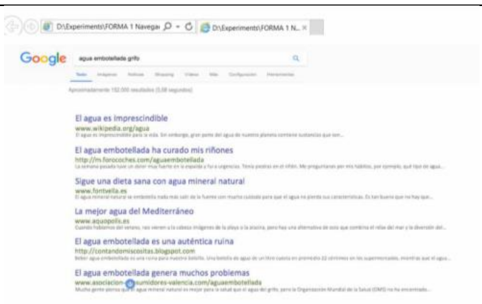
EMME #	Strategy modeled	Description	Screenshot	Time
1	Full SERP inspection	A student inspects a SERP page from top to bottom, reading at a normal pace all the page titles and some further information from the snippets. The student ends up clicking on a relevant page at the bottom of SERP after a review of the SERP titles.		69”
2	Identification of source information	A student looks at the webpage logo, reads the text once at normal pace, and finally reads the author information provided below the text.		52”
3	Deep reading of trustworthy and relevant pages	A student looks at the webpage logo (institutional page), reads the text twice at normal pace, and finally reads the author information provided below the text.		77”
4	Skim less trustworthy and irrelevant pages	A student looks at the web page logo (popular forum) and user’s information located at the left of the text, and quickly skims the text.		29”
5	Quickly abandon topically unrelated pages	A student looks at the webpage logo (commercial service unrelated to the task) and abandons the page without reading the text.		18”

Table 3

Skewness and Kurtosis for measured variables.

	Pre-test				Post-test			
	Skewness (SE)	Kurtosis (SE)	Minimum	Maximum	Skewness (SE)	Kurtosis (SE)	Minimum	Maximum
Eye-movements indices								
SERP reading times	1.46 (.30)	2.00 (.59)	22.16	423.16	1.16 (.30)	.83 (.59)	42.10	508.27
Page header dwell times	.95 (.30)	.71 (.59)	166.60	2171.23	1.25 (.30)	1.28 (.59)	158.20	3478.88
Page author information dwell times	1.30 (.30)	1.91 (.59)	216.30	4782.88	.97 (.30)	.57 (.59)	216.50	4266.88
Trustworthy pages reading times	-.02 (.30)	.28 (.59)	127.39	386.22	.27 (.30)	1.84 (.59)	94.54	457.34
Untrustworthy pages reading times	-.55 (.30)	.40 (.59)	149.08	443.19	-.59 (.30)	.29 (.59)	82.23	416.77
Essay indices								
Single paraphrases	1.90 (.30)	3.75 (.59)	0	4	2.18 (.30)	5.75 (.59)	0	6
Intra-text inference	1.19 (.30)	1.47 (.59)	0	4	.72 (.30)	-.32 (.59)	0	6
Inter-text inference	.38 (.30)	-.83 (.59)	0	3	.24 (.30)	-1.02 (.59)	0	3
Document sources	3.12 (.30)	11.42 (.59)	0	5	1.36 (.30)	.59 (.59)	0	3

Embedded sources	1.57 (.30)	2.10 (.59)	0	4	1.30 (.30)	.92 (.59)	0	3

Table 4

Means and standard deviation (in brackets) for measured variables.

	Control		EMME	
	Pre-test	Post-test	Pre-test	Post-test
Eye-movements indices				
SERP reading times	145.2 (84.5)	132.3 (59.7)	148.4 (84.9)	221.9 (117.4)
Page header dwell times	700.2 (418.3)	1072.6 (772.2)	789.9 (467.1)	1079.9 (834.1)
Page author information dwell times	1486.4 (1122.6)	1342.0 (886.4)	1105.0 (800.7)	1656.0 (1106.0)
Trustworthy pages reading times	267.5 (45.8)	258.3 (36.6)	271.1 (55.7)	285.1 (55.7)
Untrustworthy pages reading times	312.7 (61.2)	310.6 (57.2)	317.7 (65.3)	271.1 (80.6)
Essay indices				
Single paraphrases	0.7 (1.0)	0.8 (1.5)	0.7 (1.0)	0.8 (1.5)
Intra-text inference	0.7 (0.8)	0.8 (1.0)	1.1 (1.1)	1.0 (0.9)
Inter-text inference	1.1 (0.9)	1.1 (0.8)	0.8 (0.8)	0.9 (0.8)
Document sources	0.3 (0.6)	0.3 (0.7)	0.5 (1.1)	0.7 (0.9)
Embedded sources	0.5 (1.0)	0.8 (1.0)	0.7 (0.9)	0.6 (0.8)