The effect of document source trustworthiness on the evaluation and strategic use of embedded sources when reading health information online

Franco Londra & Gastón Saux

ABSTRACT

The organization of sources into layers may have an impact on the way readers evaluate conflicting documents online. Two experiments (n = 131) examined whether undergraduates use metadata from the document to evaluate the contents and embedded sources included in that document. Participants read two texts about treatments for a rare disease put forward by two neutral characters (the embedded sources). Each text was manipulated so that it was published by a trustworthy or untrustworthy document source. In Experiment 1, participants performed the task using their own criteria. In Experiment 2, they received a pre-training on how to evaluate sources. Participants used more information (cited more sources and preferred the treatment) and rated the embedded source as more trustworthy when associated to a trustworthy document, but only in Experiment 2. In conclusion, readers can strategically use multiple source layers, suggesting a networked source representation, but contingent to task
specifications

Introduction

Imagine that someone is diagnosed with a rare disease and their doctor proposes several potential treatments. Even after receiving an expert explanation of the options, this person will probably engage in internet-based research to gain more insights on the topic. It is very likely that this person will find diverse, even contradicting, viewpoints on the different treatments, especially if the condition is complex (e.g., multi-determined). As a non-expert, the reader may find it difficult to evaluate the claims and evidence put forward in each standpoint against one
another. In fact, most adults use the internet to find health information (Lim, Molenaar, Brennan, Reid, & McCaffrey, 2022), even though information on the internet usually differs from experts' recommendations (e.g., Khan, Khan, Huang, Warrian, & Gooi, 2022; Ng, Ahmed, & Zhang, 2021). The COVID-19 pandemic was a breeding ground for these situations. For example, despite that over 90% of the US population trust their doctors as reliable sources about COVID-19 information (Marlon et al., 2020) and that social media was identified as the main source of
fake news about COVID-19 (Naeem, Bhatti, & Khan, 2021), nearly half the population got vaccine's news on social media and 40% think of them as an important means to stay informed (Mitchell & Liedke, 2021). The Content-Source Integration model (CSI; Stadtler & Bromme, 2014) addresses the cognitive processes involved when readers access conflicting information. According to this model, readers need to "make validity judgments and evaluate the truth conditions of competing claims" (p. 387) to resolve conflicts. They could achieve this by asking what is true or whom to believe. Asking what is true involves using one's own knowledge to make the judgment. An experienced physician could evaluate the viability of a treatment by judging the quality of the arguments and their concordance with his/her previous knowledge. But, for laypeople, this strategy may become taxing, as the knowledge needed to adequately evaluate the information may be missing. Asking whom to believe is a kind of bypass in which source parameters are used to determine the trustworthiness of the document. Trustworthiness evaluations involve determining if someone has the competence and intentions to share reliable information (Sperber et al., 2010). In this context, the role of source evaluation is seen as a fundamental component of discourse comprehension

(e.g., Saux, Britt, Vibert, & Rouet, 2021a). Source is defined as metadata information potentially referring to a text origin, authorship, context, date, motives behind its publication, among other parameters (Braasch, Bråten, & McCrudden, 2018; Britt & Aglinskas, 2002). By integrating source information into the mental representation of what is being comprehended, a so-called *documents model* is constructed (Britt, Perfetti, Sandak, & Rouet, 1999; Perfetti, Rouet, & Britt, 1999). Three source dimensions have been suggested as critical to readers' evaluation: author position, author motivation, and media quality (Pérez et al., 2018). The author's position involves the occupation or expertise of the individual who provides information. Author's motivation refers to his/her intentions regarding the transmission of information, they could be compromised by providing accurate information or their ends could be guided by economic or ideological interests, leading to intentional bias. Last, media quality depends on the existence of editorial filters or controls regarding the information published. Personal publications, such as blogs and posts on social media, present little to no filter, whereas expert peer-reviewed publications, evaluate and select what will be published. By considering these dimensions, readers may determine which sources to believe, thus affecting their representation of the documents and how they use it afterward (e.g., Rouet et al., 2020, List et al., 2022). Nonetheless, paying attention to sources is uncommon among non-expert readers who rarely use source-based strategies to evaluate texts, even when the topic is unfamiliar to them (e.g., Britt & Aglinskas, 2002; Macedo-Rouet et al., 2019). Sourcing effects may appear only when readers are attentive to source characteristics or when participants are prompted to pay attention to them (e.g., Macedo-Rouet et al., 2019; Van Boekel, Lassonde, O'Brien, & Kendeou, 2017). When this happens, sources characteristics influence information processing during and after reading. First, differences in source trustworthiness (high vs low) have shown to increase reading times when compared to texts from sources similar in trustworthiness (Gottschling, Kammerer, & Gerjets, 2019; 2021). Texts from untrustworthy sources have also shown to increase reading times when compared with trustworthy sources (Salmerón, Delgado, & Mason, 2020). Second, memory of sources' features has been related to better learning from multiple documents (Strømsø, Bråten, & Britt, 2010). When readers are attentive to source information, they use it in written productions to justify their evaluations of what they read (Kammerer, Kalbfell, & Gerjets, 2016; Macedo-Rouet et al., 2019), agreeing more with trustworthy sources and using them to explain information discrepancy (Gottschling et al., 2019; Pérez et al., 2018).

Documentary and Embedded Sources and their Impact on Reading
Prior research worked with sources as documentary information, external to the text (e.g., Britt & Aglinskas, 2002; List, Alexander, & Stephens, 2017; Rouet, Britt, Mason, & Perfetti, 1996), or, alternatively, as textembedded information (e.g., Foy, LoCasto, Briner, & Dyar, 2017; Rouet, Le Bigot, de Pereyra, & Britt, 2016; Saux et al., 2021b). From a cognitive standpoint, the tendency is not to differentiate between documentary information and sources cited in the text, merging these different forms into a more global concept of source (for a review on this matter, see Braasch et al., 2018). However, at least formally, it makes sense to draw this distinction. Document Sources (DS hereinafter) are defined as metadata about a particular document, that is, information about where something is being published. This may include date, media, publisher, etc. A collaborative encyclopedia, a newspaper, an NGO, could be examples of DS. Originally, Bråten, Strømsø, and Andreassen (2016) used the term Primary Source to refer to this type of source. As the term primary source is commonly used to

refer to original documents or first-hand witnesses, we decided to adopt the term Document Source (Salmerón et al., 2020) to avoid potential misunderstandings. Conversely, Embedded Sources (ES hereinafter) have been defined as sources cited within a document: a physician who explains a treatment, a passerby that gives his opinion, a politician who presents his plans (Bråten et al., 2016). Note that not only characters could be ES, a newspaper referred to in a text as the source of a statement could be thought as an ES, too. Therefore, the distinction between DS and ES is contextual in nature, as the same source can act as DS or ES in different situations. The organization of sources into layers may have an impact on the way readers evaluate documents, although evidence about this specific issue is still scarce. Strømsø, Bråten, Britt, and Ferguson (2013) asked 18 undergraduates to think-aloud while reading documents about a controversial health issue to advise a friend. After reading, participants were asked to write a brief essay about the topic. Analyses of think-aloud data showed that participants paid attention to source information, from both DS and ES. In the essays, DS were cited more often than ES. In a later study (Strømsø & Bråten, 2014) used a similar design on a larger sample. It corroborated previous findings: students paid attention to both kinds of sources, although they rarely used them for evaluative purposes. Again, DS were cited more often than ES. In 2016, Bråten et al. asked college students to read two texts, each discussing a different controversial issue (i.e., artificial sweeteners' and cell phones' effect on health), to decide if they would change their habits and to answer some questions. After reading, participants were asked to state the conclusion of each text and to provide the sources of these, first spontaneously and then cued. Memory was low for DS and lower for ES, even when cued. Authors argued that having read only one text per topic could have favored their lack of attention to source information. Salmerón, Strømsø, Kammerer, Stadtler, and van den Broek (2018) asked their participants to read either real (i.e., on textbooks, newspapers, magazines) or print-out versions (i.e., same texts but printed on white A4 paper sheets) of a set of conflicting documents on climate change and requested them to write a report describing and evaluating the different perspectives. In general, few sources were cited in the writing task. Interestingly, participants made more references to DS after reading real texts, but ES references did not change between text types. Reading real texts may have turned DS more distinctive, as physical cues may improve memory for DS information without affecting ES distinctiveness (Salmerón et al., 2018). More recently, Salmerón et al. (2020) asked their participants to read and work with texts about controversial socioscientific issues (i.e., climate change or GMO food) as part of an intervention study. It involved showing participants the gaze patterns of "good readers" in an attempt for them to adopt their strategies. First, they found that most people did not include any citation on their productions, neither from DS nor ES. Second, after the intervention, DS citations increased, but ES citations did not. Additionally, they found that the intervention modified reading times. Before it, the participants spent more time reading pages from untrustworthy sources. After the intervention, this difference disappeared, resulting in similar reading times for trustworthy and untrustworthy pages, suggesting a quicker discard of the latter (Salmerón et al., 2020). To sum up, earlier works using think-aloud methodology found that people paid attention to both kinds of sources and that they included them in essays (Strømsø et al., 2013; Strømsø & Bråten, 2014). More recent studies found that memory for DS and ES was very low (Bråten et al., 2016) and that they were scarcely cited on written productions (Salmerón et al., 2018; 2020). Across most of the studies DS were cited more when compared to ES. Still, to the best of our knowledge, the relationship between different sources' layers and their use in evaluative situations remains unclear. Strømsø et al. (2013) proposed that there could

be multiple ways in which different layers of source information could be represented: first, readers could see the ES merely as a character in the story; second, readers could represent the ES as a source of information disregarding the DS; and third, a more expert way would be to contextualize the ES as a source in the context of their DS. However, these alternative ideas have not been yet tested empirically in a direct fashion.

The Present Study

Our general goal was to explore the role of sources' layers in the evaluation and adoption of medical information. Specifically, we set to examine if the trustworthiness of Documents Sources (DS) impacted the trustworthiness evaluation of Embedded Sources (ES), the sources' mentions in essays, the adoption of the information they proposed, and
reading times. We conducted two experiments in which undergraduate psychology students read two texts on treatments for a rare genetic condition (i.e., cystic fibrosis) and were asked to write an essay indicating which treatment was the best. Each text included information about their DS and an ES supporting a specific treatment. DS were manipulated into being trustworthy or untrustworthy, whereas the ES remained neutral (i.e., mid-level of trustworthiness). After completing their essays, participants were asked to rate the trustworthiness of every source. The main difference between experiments was a pre-training, added to Experiment 2, about the relevance of source information in evaluative situations.

Experiment 1
The objective was to determine if DS trustworthiness influenced ES trustworthiness evaluation when participants read to make decisions about little known health topics. Our first hypothesis was that reading times (i.e., time spent reading a section of the text) would differ as a function of the trustworthiness of the DS (H1). In particular we were interested in three critical sections: DS information, ES information, and the description of the treatment. Second, we expected that written productions would include more trustworthy DS than untrustworthy DS (H2a) and more ES presented in the context of trustworthy DS when compared to those presented with untrustworthy DS (H2b). Third, we expected that the treatment proposed by an ES in the context of a trustworthy DS would be chosen as the best more often, when compared to the treatment proposed by an ES under an untrustworthy DS (H3). Finally, we hypothesized that an ES presented in the context of a trustworthy DS would be evaluated as more trustworthy than those in the context of an untrustworthy DS (H4).

Participants

Twenty-seven undergraduate students volunteered as participants (Age $M = 20$, $SD = 2.53$). All of them were first to third-year psychology students from a large South American university. Participants were native Spanish speakers and signed an informed consent. Participants self-reported knowledge about cystic fibrosis was generally low, $M = 2.44$ $SD = 2.08$, out of a maximum of 10. The original sample was of twenty-nine, but two participants had to be removed due to faulty comprehension of the instructions.

Materials and Methods
The main materials were two texts about treatments for a rare genetic disease. Each text was composed of 6 sections. Section 1 gave information about the DS, which could be either trustworthy (i.e., proceedings of a meeting) or untrustworthy (i.e., periodic publication of a pseudoscientific society). Section 2 provided general information about the disease. Section 3 introduced the ES (an individual described by

his/her name, occupation, and workplace) who proposed a treatment. Sections 4 and 5 presented the treatment plus a description of it. Section 6 served as a closing by providing general epidemiological information. Section's length varied, but comparisons were always between the same sections across conditions. Total length was similar for both texts (160 and 175 words). The materials were implemented on Visual Basics for Applications (Balena & Fawcette, 1999) to be presented on screen in a self-administered manner. The software included the texts, written instructions, and response fields, which were presented as the study unfolded (see Procedure for further details). The experiment presented a single within-subject manipulation. Each participant had to read both texts. One with a trustworthy DS and the other with an untrustworthy DS. The associations between DS trustworthiness and text version were counterbalanced. The inclusion of the neutral ES (section 3) was also counterbalanced, so that each ES could be read in the context of either text and under trustworthy or untrustworthy DS. In order to access the texts, participants had to press one of two buttons on the screen (labelled "Text 1" and "Text 2"), the text assigned to each button was randomized by the experimental software in each trial at the beginning of the task.

Three sets of dependent variables were collected. First, reading times were recorded in seconds via an adaptation of the moving-window technique (Just, Carpenter, & Woolley, 1982). All sections were blurred and participants had to click on each section to make it readable. One section was readable at a time, when a new section was clicked the latter became blurred again. Reading times of an area were defined as the lapse between the click to enable a section and the following click. In case of multiple visits to the same area, time lapses were merged into a total value.

Second, after reading, participants were asked to write an essay stating which the best treatment for the disease was. Three indicators were extracted from the analysis of the essays: treatment selection, DS mention, and ES mention. As our interest was to observe if people selected the treatment proposed by the trustworthy DS we coded it as follows: if the essay stated that the best treatment was the one proposed in the trustworthy condition (i.e., by a trustworthy DS) it was coded as 1, otherwise, it was coded as 0. Responses coded as 0 could include: the selection of the treatment in the untrustworthy condition, stating that both treatments were good or that neither was worth recommending. DS and ES mentions were coded as 1 if any direct feature of the source (name, occupation, workplace) or indirect reference (e.g., the person who gave his/her opinion in the first text) was mentioned at least once and 0 if it was absent.

Third, after finishing the essay participants were asked to rate the trustworthiness of each source, document and embedded, on a 0 to 10 scale. Items were presented in random order. At this moment participants could not consult their essays nor the texts.

Preliminary Testing and Manipulation Check

Texts were constructed from online information about cystic fibrosis and its treatments. In a first step, 4 definitions of the disease and 4 treatments were adapted as independent paragraphs of similar length, and a list of 14 ES was created. These sources varied in expertise (e.g., pulmonologist, nurse, teacher) and first-hand knowledge (e.g., therapeutic companion, scholar). Closing paragraphs were not relevant for the manipulation and

therefore were not pretested. Second, to pretest these materials, 120 first to third-year university students (age $M = 20.64$, $SD = 4.05$) volunteered to rate their trustworthiness. After an introduction, they were presented with the 4 definitions of the disease, the 4 treatments, and the 14 ES and were asked to rate their trustworthiness on a 10 points scale. Before rating the trustworthiness of the ES, participants were prompted to rate them considering they would give information about cystic fibrosis. Based on participants' trustworthiness ratings, two equivalent versions were constructed. Each version consisted of one definition, one treatment, and one ES. Each section was similar in perceived trustworthiness to its pair, $p > .05$. Table 1 presents item comparisons.

Finally, to establish whether high and low DS were perceived as expected, the pretesting sample ($n = 120$), also rated the trustworthiness of a list of 13 DS varying in trustworthiness, specifically in terms of pertinence and media-quality. Two DS with extreme ratings were selected:

Table 1

Descriptive and t-tests for the evaluation of the trustworthiness of texts' sections (preliminary testing)

	Text A	Text B	t (df)	p-value
	M (SD)	M (SD)		
Definition	5.12(2.33)	4.98(2.35)	.51(118)	.62
Embedded Source	5.88(2.93)	5.93(2.43)	-.151(115)	.88
Treatment & treatment description	4.75(2.24)	4.67(2.31)	.3(107)	.77

Note: Text A and B stands for the two versions constructed from the selected information.

the proceedings of a scientific meeting, $M = 6.84$, $SD = 2.55$; and a magazine by a seudoscientific society, $M = 1.81$, $SD = 1.58$; t (117) = -18.76, $p < .001$. The final texts were constructed with the selected definitions and treatments, presenting similar length (160 vs 175 words). To ensure that both were similar in complexity, two analyses were conducted. First, they were analyzed with the Expository Text Analysis Tool (Vidal-Abarca et al., 2002), yielding similar results of distance and number of connections between semantic nodes, $p > .53$. Second, both texts showed "somewhat difficult" readability level (57.53 vs 59.12), as informed by the Fernández-Huerta index (Fernández Huerta, 1959).

Procedure

The experiment was conducted in a computer lab on two groups of approximately 15 participants. First, they were handed out the informed consent. Second, they were assisted to start the software. Third, each participant answered socio-demographic questions and rated their knowledge about cystic fibrosis. Fourth, the moving-windows technique was explained and the participants read a practice text about global warming. Fifth, instructions were read out loud: "Imagine you had been learning about

cystic fibrosis, a rare genetic disease. As homework, you were asked to read two texts on this subject and to write an essay stating

what the best treatment for the disease is. Imagine you searched on the internet and found the following texts. Read them carefully as you'll have to work with them. It's important to be aware of where you are getting the information and who presents it". Any doubt regarding the instructions was cleared out and a research assistant was available for any further questions. Last, participants completed the task by reading both texts, writing their essays, and rating the trustworthiness of each DS and ES. While writing their essays texts remained available. The full

procedure lasted approximately one hour.

Statistical Analysis

Analyses were conducted in R 4.1.0 (R Core Team, 2021) with the addition of the package lme4 (Bates, Mächler, Bolker, & Walker, 2015) and lmeTest (Kuznetsova, Brockhoff, & Christensen, 2017). To explore the potential effect of DS trustworthiness on reading times (H1) we fit a linear mixed model for the general reading time (H1a) and one for each section of interest (H1b). Given the non-normal distribution of reading times, natural logarithmic transformations were applied. Transformed reading times were set as the response variable, the condition (i.e., DS trustworthiness) and prior knowledge were added as fixed factors, and random intercepts for each participant were added. Following Berger and Kiefer (2021), outliers that exceeded twice the standard deviation before transformation were excluded from the analysis resulting in a mean exclusion of 5.13%, $SD = 1.57$. For H2a and H2b (i.e., influence of DS trustworthiness on sources mentions), and H3 (i.e., influence of DS trustworthiness on treatments selection) chi square tests were applied. Note that mention was coded as a binomial response, regardless of their occurrences. To test H4 (i.e., influence of DS trustworthiness on ES trustworthiness rating) we fit a linear mixed model. ES trustworthiness rating was entered as the response, and the condition and prior knowledge were added as fixed factors. Additionally, random intercepts were added for each participant. For the manipulation check we fit linear mixed models with DS trustworthiness rating as response and the condition (i.e., DS trustworthiness) and prior knowledge as fixed factors. Again, we add random intercepts

for each participant.

Results

Reading and Essays Productions

First, we assessed the potential effect of DS trustworthiness on reading times. Reading times were analyzed globally and particularly for each section of interest. First, descriptive analyses of reading times are presented in Table 2. Participants spent between 33 seconds and 2:17 minutes to read each text. As expected, they spent most of the time reading the longest sections (definition of the disease and description of the treatment). An inspection of the means by condition showed a tendency, as most sections from untrustworthy documents were read for longer, except for section 5.

Table 2

Experiment 1: Mean, standard deviation and regression parameters for reading times

	Trustworthy DS	Untrustworthy DS	Regression parameters
			Document Source trustworthiness

	Mean	SD	Mean	SD	Est.	B	t	P
Total	82.88	27.44	83.36	27.19	4.39	-0.01	-0.13	.89
Section 1: DS	9.52	4.22	10.4	5.92	2.34	-0.03	-0.21	.84
Section 3: ES	6.32	3.67	6.68	3.2	1.76	-0.11	-0.71	.49
Section 5: treatment description	14.88	5.28	14.52	4.81	2.59	0.01	0.07	.94

Note: Table presents data measured in seconds. Calculations were made after the trimming process. Reference level for DS trustworthiness is "Untrustworthy DS". Est. stands for Estimate. Prior knowledge, also included as a fixed factor, did not yield significant results.

Nonetheless, the differences were small. As expected by the small differences, the models confirmed the lack of effect of DS trustworthiness on reading times (Table 2). Previous knowledge also failed to predict both total and specific reading times, all $p > .13$. Analyses of the remaining areas can be consulted in the Online Appendix, supplementary materials. Second, we assessed whether our participants used source information in their essays as a function of the manipulation. Trustworthy DS were mentioned in 57.14% of them while untrustworthy DS were mentioned in 42.85%. ES in the trustworthy condition were mentioned in 67.85% of them and ES in the untrustworthy one were mentioned in 50%. A significant difference as a function of our manipulation was found for trustworthy ES. Chi-square tests showed that trustworthy ES were effectively more cited than not, $X2 (1) = 4.81$, $p = .034$. On the other hand, for both kinds of DS, and ES read in an untrustworthy context chi-square tests showed non-significant results, all $p > .336$. Third, we analyzed whether the treatment proposed in the trustworthy condition was selected as the best more often. Descriptive analyses showed that the treatment by the trustworthy source was selected as the best in nearly half of the essays (14) while the other half selected the treatment by the untrustworthy source or neither of them (13). As expected, this difference was non-significant, $X2 (1) = .037$ $p = .847$.

Trustworthiness Evaluations

Finally, we examined the post-reading trustworthiness ratings for the DS (manipulation check) and the ES. A first analysis of ES mean ratings across conditions showed similar values, trustworthy DS condition: $M = 5.93$, $SD = 2.46$; untrustworthy DS condition: $M = 5.48$, $SD = 2.03$. To assess this difference, we fit the proposed linear mixed model. None of the predictors explained differences in the rating of the ES; DS condition: $B = .44$, $t = 0.72$, $p = .47$; prior knowledge: $B = .12$, $t = 0.79$, $p = .43$. Regarding DS (i.e., the manipulation check), the scores presented considerable variation, ranging from very low scores (1 or 2) to the maximum score (10), both for the trustworthy ($M = 7.48$, $SD = 2.21$) and the untrustworthy DS ($M = 4.44$, $SD = 2.62$). The model showed that our manipulation had an impact on the trustworthiness ratings of the DS, $B = 3.04$, $t = 5.44$, $p < .001$. Prior knowledge did not predict the ratings, $B = 0.08$, $t = 0.45$, $p < .658$.

Discussion

In this experiment, we asked participants to read two texts about a rare genetic disease in order to write an essay stating the best treatment. By doing so, we sought to examine whether Document Sources trustworthiness would influence reading times of critical areas of the documents (H1), Document and Embedded Sources mentions in

the essays (H2a and b, respectively), the preference for a treatment (i.e., treatment selection task; H3), and modify the trustworthiness ratings of Embedded Sources (H4). Our results primarily support H2b, the remaining data showed the expected directing but the differences across conditions were statistically non-significant. Regarding reading times (H1) participants spent similar lapses reading both the trustworthy and the untrustworthy documents. This result is unexpected because prior research has reported online effects as a function of differences in source trustworthiness during reading (e.g., Gottschling et al., 2021; Salmerón et al., 2020). However, the online effect sometimes is less clear. For example, Van Boekel et al. (2017) observed no effects on reading times if their participants were not prompted to pay attention to source information. Although our study included a source evaluation prompt in the instructions, it may have not been strong enough to induce a source-based reading strategy. In this line, the lack of effect of our manipulation could be explained by little attention to the cues that set the sources apart, or a lack of strategic allocation of attention based on these cues. Another explanation could be that participants had to click on a specific area to read it. This could havemodified their natural reading behavior, for example discouraging revisits to previous areas. Regarding sources mentions in essays (H2a and b) readers effectively included more references to the ES that were read in the context of a trustworthy DS. Given that the topic was unfamiliar to the participants, it was expected that they turn to the sources when evaluating the texts(Stadtler & Bromme, 2014). Interestingly, DS mentions were unmodified by the manipulation, pointing out at least some kind of discrimination between ES and DS. Designating a source as untrustworthy depends on knowledge about the source. The preliminary study (n = 120) and the manipulation check (at the end of the main study) were our strategies to ensure that participants were perceiving the sources as intended. Nonetheless, nearly 15% of the participants of the experiment rated their trustworthiness as opposite to what we expected, giving the trustworthy DS lower scores than the ones of the untrustworthy DS. We interpret that these participants may have used other criteria to evaluate trustworthiness. In fact, people can rely on a wide variety of parameters when evaluating the trustworthiness of sources (Britt & Aglinskas, 2002) some more proficient than others (e.g., Hargittai, Fullerton, Menchen-Trevino, & Thomas, 2010; Wineburg & McGrew, 2019). These parameters do not necessarily coincide with a normative list of dimensions. For example, individual differences in epistemic beliefs, prior knowledge or motivation can also affect the way people evaluate sources (Anmarkrud, Bråten, Florit, & Mason, 2022; Barzilai & Eshet-Alkalai, 2015; List & Alexander, 2017). Following this line of thought, we retested H4 to evaluate whether the *perceived trustworthiness* of the DS (as evidenced by the ratings provided for the manipulation check at the end of the experiment) had any effect on ES evaluations. To do so, we fit a linear mixed model with the ratings of the DS provided in the manipulation check as a fixed continuous predictor (instead of the manipulation based on the preliminary testing) and ES ratings as outcome. A significant effect of DS ratings was found, indicating that higher scores for the DS were associated with higher scores for the ES, B = 1.31, t = 2.14, p = .04. Considering these late analyses, we decided to design a second experiment. Experiment 2 replicated the structure and hypotheses of the first study but included a pre-training on source evaluation that was held before the main tasks. This was done to ensure that the trustworthiness of the DS was perceived as intended and similarly across participants.

Experiment 2
Acknowledging the limitations of Experiment 1, a pre-training on source evaluation when reading online information was held before the main activities. Students were

explicitly instructed to complete the tasks with the pre-training in mind. Previous research has found that given explicit instructions (e.g., Stadtler, Paul, Globoschütz, & Bromme, 2015) and training on the relevance of specific source dimensions in evaluative situations (e.g., Pérez et al., 2018) students could improve their attention and strategic use of available source features. The objective of this experiment was to determine whether, under these conditions, DS trustworthiness would have any effect on reading times, on the selection of the treatments provided, on sources references in their essays, and ES trustworthiness evaluation. We worked with the same hypotheses of Experiment 1.

Participants
104 students from a psychology course of a large South American university (age $M =$ 25.01; $SD = 8.26$; 86 females 17 males and 1 other) participated as volunteers. Prior to participating, they had to sign an informed consent. Every participant was a native Spanish speaker. The original sample was composed of 107 participants but three of them had to be removed because they did not comply with the task. Self-reported prior knowledge about the topic of the documents (cystic fibrosis) was low, $M = 1.54$, $SD = 1.70$.

Materials and Methods
The design was like that of Experiment 1, except that it was adapted to be administered online (data from Experiment 2 was collected during the mandatory COVID-19 lockdown). A web application was developed using PHP and JavaScript. Besides minor changes (i.e., font, button alignment, interface colors, etc.), the main difference with Experiment 1 was related to the implementation of the moving-windows technique. Instead of clicking in the blurred sections to read them, participants in Experiment 2 simply had to hover the mouse above a section to unmask it. As in the offline version, only one area was visible at a time. As in Experiment 1 the original intention was to record chronometric data during reading. However, due to compatibility issues between the application and some of the participants' software reading time data from 45 participants (43.3% of the sample) were not collected. In light of this limitation, we decided not to report chronometric data for Experiment 2 in the Results section. Nonetheless, we will review the results of the available sub sample in the Discussion.

Pre-training on Source Evaluation
As stated before, Experiment 2 added a pre-training on the relevance of key source features when evaluating the trustworthiness of a document. The pre-training was adapted from Pérez et al. (2018) and consisted of a 25 minute group reflection about how and when to assess sources when reading online. First, the participants were presented with the relevance of considering the characteristics of the author (who says what), their motives (why the author says it), and media quality (where it is published) when reading information online. Second, several examples and vignettes were presented, and a group discussion was promoted to reflect on the explanations of the first part. These examples included DS and ES, but this distinction was not pointed out nor was the manipulation; that is, using one of the sources (e.g., the DS) to weigh the evaluation of the other source (e.g., the ES). The full procedure was conducted online in synchronous sessions in groups of about 30. The first author coordinated the whole procedure via a video conference APP (Jitsi meet). This included the pre training,

reading out loud the instruction to ensure the comprehension of the task and to clarify doubts if needed. First, participants signed the informed consent and received the pre-training (30 min). Then, they accessed the experimental task via a link provided by the coordinator. As in Experiment 1, the task was designed to be self-administered. The pre-training and the experimental tasks took around one hour.

Statistical Analysis

Analyses in Experiment 2 followed the same strategy of Experiment 1. To evaluate the potential association between DS trustworthiness and sources' mentions (H2a and H2b), and treatment selection in the essays (H3), chi-square tests were applied. To test influence of DS trustworthiness on ES evaluation (H4) and the manipulation check we fit linear mixed models with random intercepts for participants. The first model (H4) included ES trustworthiness rating as the response and DS trustworthiness and prior knowledge as fixed factors. The second model (manipulation check), included DS trustworthiness ratings as response and DS trustworthiness and prior knowledge as fixed factors.

Results

Essay Production

First, regarding sources' citations, trustworthy DS were mentioned on 53.8% of the essays, while untrustworthy DS were mentioned on 28.84% of them. ES under trustworthy DS were mentioned in 70.2% of the essays while ES under untrustworthy DS were mentioned in 49% of them. H2 predicted that source inclusion in the essays would vary as a function of the trustworthiness of the DS. Trustworthy DS were mentioned in nearly half of the essays (53.8%), $X2(1) = .615$, $p = .433$. In turn, untrustworthy DS were more likely to be excluded from the essay (28.84%), $X2(1) = 18.62$, $p < .001$. Regarding ES, participants included them in the essays significantly more often when they were read in trustworthy documents (71.15%), $X2(1) = 16.96$, $p < .001$, but not in untrustworthy documents (45.19%), $X2(1) = .038$, $p = .845$. Second, regarding treatment selection, 63.5% of the participants selected the treatment by the trustworthy DS, 26% by the untrustworthy DS, and 10.6% declared that neither was worth recommending. Recommending the treatment by the trustworthy DS was significantly higher than the other two response types, $X2(1) = 7.54$, $p = .006$.

Trustworthiness Evaluations

Last, we tested if DS trustworthiness increased the trustworthiness ratings of the ES. A linear mixed model was fitted with ES trustworthiness ratings as the response. A significant main effect of our manipulation was found, $B = 1.33$, $t = 3.32$, $p = .001$. ES read in the trustworthy condition received higher scores, $M = 5.91$ $SD = 2.75$, than those in the untrustworthy one, $M = 4.59$ $SD = 3.03$. Prior knowledge did not predict the ES ratings, $B = 0.18$, $t = 1.58$, $p = .117$. Regarding DS (i.e., manipulation check), mean trustworthiness ratings of trustworthy DS were considerably higher than those from untrustworthy DS, $M = 8.14$, $SD = 2.61$ vs $M = 2.84$, $SD = 2.65$. While controlling for prior knowledge, which was not a significant predictor, $B = -0.134$, $t = -1.23$, $p = .22$, the model showed that the trustworthy DS were rated as more trustworthy than the untrustworthy DS, $B = 5.31$, $t = 14.83$, $p < .001$.

Discussion

In our second experiment, we set to evaluate if eliciting strategic reading with a brief training the trustworthiness of Document Sources would impact the attributed trustworthiness to the Embedded Sources, the sources' citation on written productions, and the selection of the treatment for a disease. As expected, the pre-training was effective, resulting in a more consistent and less variable response among participants and in a clear difference in DS assessments as a function of the manipulation. A first set of predictions related to written productions. We expected that the trustworthiness of the DS would influence reading times (H1). As discussed in the design, a failure in the recording of the chronometric data affected 43.3% of the sample, so we decided not to include these analyses in the results. However, mixed modeling on the reading times of the remaining 56.7% of the sample (GLMM with Gamma distribution and identity link, Lo & Andrews, 2015) revealed an effect of the manipulation of the DS on the time spent reading the DS and the treatment description. On average, participants took 2.11 seconds less, *CI* 95% [-3.53, -0.69], to read information about a trustworthy DS, $t = -2.94$, $p = .003$, and 1.54 seconds less, *CI* 95% [-2.98, -0.1], when treatment descriptions were read in the context of a trustworthy DS, $t = -2.12$ $p = .034$. Although they should be taken with caution, these results are in line with H1 as the trustworthiness of DS affected the reading times of two of the three critical sections (DS description and treatment description). They are also consistent with prior findings suggesting that differences in sources trustworthiness alter reading times (e.g., Gottschling et al., 2021, Kammerer et al., 2016). Specifically, we found that a trustworthy DS lead to shorter reading times. This could be explained by validation processes (Richter & Maier, 2018). Finding a trustworthy DS could lead to a less vigilant stance regarding not only the source but the content included in the document (i.e., the treatment). However, spending extra time reading an untrustworthy source could lead to a backlash effect, increasing the effects of exposure to misinformation such as confusion (i.e., mistrusting one's own knowledge when confronted with contradicting assertions), doubt (i.e., revising one's own knowledge even when true), and reliance (i.e., using misinformation to reach goals; Rapp & Salovich, 2018; Hassan & Barber, 2021). Please note though, that future research would be needed to support this assumption; for example, we cannot discard that, if more texts were available, readers would outright reject texts from untrustworthy sources (e.g., Salmerón et al., 2020). Second, regarding written productions, we expected that readers would refer more to the trustworthy DS (H2a) and ES (H2b) when compared with their untrustworthy counterparts. In line with these predictions, untrustworthy DS were excluded from nearly three-quarters of the productions, while trustworthy DS were included nearly half of the times. The opposite was found for ES references: participants included reference to them in almost three-quarters of their essays when read in trustworthy documents while no difference was found when they were read in untrustworthy contexts. In addition, the participants turned primarily to the treatment proposed in the trustworthy DS, with only 26% siding with the untrustworthy DS and 10.6% refraining from recommending any treatment. Overall, data from the written production provides evidence that the participants not only added source information, but they also strategically used this information by deciding to include meta-textual references and being more prone to recommend the treatment associated with the trustworthy DS. Compared to Experiment 1, in which only some of the effects expected for essay productions were significant, these results are more consistent with H2 and H3. This is probably due to the pre-training in source evaluation included in Experiment 2. Indeed, explicitly prompting to attend to and/or evaluate sources, either from brief instructions or intervention trainings, tends to improve performance, at least on immediate and trained-like tasks (e.g., Brante & Strømsø, 2018; Macedo-Rouet et al.,

2019). Also, it should be noted that the pre-training did not elaborate on the integration of different source layers (DS and ES). Third, regarding trustworthiness ratings we proposed that ES would be perceived as more trustworthy when included in a trustworthy rather than in an untrustworthy DS (H4). As predicted, the DS manipulation changed by 1.33 (on a 0-10 scale) the trustworthiness ratings of the ES indicating that participants took into consideration the DS when evaluating the ES. Strømsø et al. (2013) proposed three ways in which different source layers could be represented: readers could represent ES as characters in the story instead of as information sources, they could think of the ES as sources of information disregarding the DS, or they could think about the ES in the context of the DS (considered as the more proficient alternative). As the manipulation of the DS influenced the perceived trustworthiness of the ES, the possibility that the ES was represented detached from the DS must be ruled out. Therefore, ES may have been represented as characters in the story or as a second layer of source information. Nonetheless, the fact that ES linked to trustworthy DS were included more often in the essays suggests that these were used as elements to validate the argumentation and not mere characters. Furthermore, the fact that references to ES were more frequent in the essays than to the DS put forward the idea that the ES was represented as the main source of information while the DS modulated its trustworthiness.

General Discussion
The ability to integrate and evaluate written content with its sources is considered a key component of reading literacy. This is even more so in contexts where multiplicity of perspectives and reliability are relevant to the reading product, such as when reading on the internet (Alexander, 2020; OECD, 2021; Salmerón et al., 2018). The construction of rich inter-textual representations, a fundamental component of proficient reading, supposes that the reader will establish connections between multiple sources, and not only between sources and contents (Britt & Rouet, 2012). In this work we examined an aspect of that claim, namely, that sources external to the text (Document Sources) could influence the perceived reliability of the sources embedded in the text (Strømsø et al., 2013). Overall, undergraduate students were able to use Document Sources (DS) to evaluate, not only the Embedded Sources (ES) but also the information provided in the documents. These results were clearer in Experiment 2 in which a pre-training on source evaluation was included. Taking together the patterns observed in both experiments, two main conclusions can be derived from our results. First, the fact that the trustworthiness of the Document Source changed how readers perceived the Embedded Source indicates that the participants associated one source to the other. Initially, we assumed a layered representational organization of the sources (Strømsø et al., 2013). The idea of layers may imply directionality, so that the upper level can modify the lower level but not the other way around. Our data are not sufficient to support this claim. However, they provide enough information to assume that readers connected the sources in *some* way. In other terms, our results are not sufficient to claim that readers perceived document and embedded sources as such, that is, as distinct types of sources organized into a layered representation. Still, they point toward the claim that readers used one to evaluate the other. Following seminal contributions on multiple document modeling, readers can construct a so-called source model when confronted with multiple views (an interlinked representation of rhetorical relationships between sources and their features; Britt et al., 1999; Perfetti et al., 1999). If used strategically, this source model could be used to determine what or whom to trust (Rouet et al., 2020; Stadtler & Bromme, 2014). Second, providing participants with

a pre-training on how to read strategically by paying attention to critical dimensions of sources (Experiment 2) rendered the influence of the Document Source on the Embedded Source more transparent. Note that in Experiment 1 this effect was observed after considering the scores from the manipulation check rather than the original manipulation. In other words, participants were using idiosyncratic criteria to evaluate the documentary source, and those criteria explained the trustworthiness level perceived in the Embedded Source (see Discussion of Experiment 1). It could be the case that readers use diverse cues when determining what to trust. In fact, even though previous work has shown that students are aware and make use of many source features when evaluating what they read (Britt & Aglinskas, 2002) they may also lean to more shallow cues such as the looks of a website (Wineburg & McGrew, 2019), personal preferences of search engines (Hargittai et al., 2010), number of citations (Rouet et al., 2018), and be influenced by their beliefs about knowledge (Barzilai & Eshet-Alkalai, 2015), among other factors. However, after the pre-training (Experiment 2) students perceived the Document Sources as expected, indicating that the evaluation criteria were more homogeneous. It should be noted that participants were never prompted to use one source to weigh the other, neither in Experiment 1 nor 2. Therefore, we must assume that the connection between the two sources was made spontaneously in both studies.

Advanced readers may not need to learn from scratch about source trustworthiness and how to use this piece of knowledge purposefully. In this research, a short discussion on critical dimensions on how to evaluate sources was sufficient to observe an effect. Digital natives tend to self-report high confidence in their ability to evaluate information when using social networks and collaborative sites (Helsper & Eynon, 2013). However, this self-confidence does not always correlate with their performance (Mahmood, 2016). By narrowing the gap between the reader's subjective efficacy and their actual performance, our results underline the potential of interventions in educational settings to fine-grain evaluative skills on undergraduates. Furthermore, the heterogeneity of criteria for evaluating sources observed in Experiment 1 and the impact of pre-training in Experiment 2 suggest that continued instructions in online critical reading, and in particular in documentary evaluation, are warranted, even in advanced education levels and, perhaps, throughout a lifelong perspective (Alexander, 2020).

Limitations
This work does not come without limitations. First, the population under study were psychology students reading about a rare genetic disease. The results should be generalized carefully to other populations and/or topics. Second, the sample of the first experiment was relatively low and this may have compromised statistical power. Third, with the objective of promoting the critical evaluation of the texts, the materials were fairly long. Future studies should complement this study with shorter texts that allow for repeated trials, increasing the measurements' reliability. Fourth, the selection of the embedded sources in both experiments followed an empirical criterion: that the target population attributed them similar and mid-range trustworthiness scores (see Preliminary testing and manipulation check in Experiment 1 for a detailed description). However, the expertise of the selected ES (an administrative employee in a health research institute and a nurse) may not have been totally equivalent from a normative point of view. Thus, we cannot rule out that a pair that was equivalent in this respect would have increased the effects observed in these experiments. Fifth, we cannot discard a potential influence of the home-based implementation of Experiment 2

due to lockdown. For example, the incompatibilities with the participants' computers resulted in a considerable amount of data loss (see Experiment 2 Materials and design). Also, even though data collection was made in synchronous sessions together with the researcher, the participants may have been more distracted or may have engaged differently with the task, as compared with a lab or classroom setting. Research on digital text comprehension has recently suggested that, although power decreases and individual differences become more relevant, some experimental effects associated to digital reading found in lab (e.g., the distracting effect of irrelevant information in hypertexts) are also found at home settings (Burin, González, Martínez, & Marrujo, 2021). However, we cannot rule out a potential variation of the results as a function of the reading context (see e.g., Britt, Durik, & Rouet, 2022). As a counterpart, home is a more frequent environment for health-related queries on the internet than a classroom. Therefore, although less controlled, it could be considered a more ecological context than experiments traditionally conducted in laboratories or educational settings. Perhaps conducting such tasks at home would more accurately reflect how people read on the Internet (Britt et al., 2018). Future research should focus more on how to elicit discussions about source evaluations in educational settings that mirror everyday contexts.

Conclusion

The contribution of this paper can be summed up in three main points. It extends prior work on the construction of source-to-source links in the reader's mind (Britt, Rouet, & Braasch, 2012; Britt & Rouet, 2020; Saux et al., 2021b), by showing that a connection can be established even between sources with different formal attributes (a set of metadata external to the text, and a person providing information in the text).

Also, this study provides evidence that readers do construct links between multiple sources even without being trained and with relative ease, as long as the naïve (non-expert) criteria they draw on are taken into consideration. Finally, a brief and focused pre-training was sufficient to modify such criteria, showing the task-dependent and dynamic nature of the use of sources when reading on the Internet about health issues to make informed decisions.

Acknowledgments

References

Alexander, P. (2020). What research has revealed about readers' struggles with comprehension in the digital age: Moving beyond the phonics versus whole language debate. *Reading Research Quarterly*, *55*(S1), S89–S97. doi:10.1002/rrq.331

Anmarkrud, Ø., Bråten, I., Florit, E., & Mason, L. (2022). The role of individual differences in sourcing: A systematic review. *Educational Psychology Review*, *34*, 749–792. doi:10.1007/s10648-021-09640-7
Balena, F., & Fawcette, J. (1999). *Programming Microsoft Visual Basic 6.0* (Vol. 1). Washington, DC: Microsoft Press.

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

103. doi:10.1016/j.learninstruc.2014.12.003

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1), 1–48. doi:10.18637/jss.v067.i01

Berger, A., & Kiefer, M. (2021). Comparison of different response time outlier exclusion methods: A simulation study. *Frontiers in Psychology*, *12*, 675558. doi:10.3389/fpsyg.2021.675558

Braasch, J. L., Bråten, I., & McCrudden, M. T. (2018). Introduction to research on multiple source use. In J. L. G. Braasch, I. Bråten, & M. T. McCrudden (Eds.), *Handbook of multiple source use* (pp. 1–13). New York, NY: Routledge. doi:10.4324/9781315627496

Brante, E. W., & Strømsø, H. I. (2018). Sourcing in text comprehension: A review of interventions targeting sourcing skills. *Educational Psychology Review*, *30*(3), 773–799. doi:10.1007/s10648-017-9421-7

Bråten, I., Strømsø, H. I., & Andreassen, R. (2016). Sourcing in professional education: Do text factors make any difference? *Reading and Writing*, *29*(8), 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*(4), 485–522. doi:10.1207/S1532690XCI2004_2

Britt, M. A., & Rouet, J.-F. (2012). Learning with multiple documents: Component skills and their acquisition. In J. R. Kirby & M. J. Lawson (Eds.), *Enhancing the quality of learning: Dispositions, instruction, and learning processes* (pp. 276–314). Cambridge, UK: Cambridge University Press. doi:10.1017/CBO9781139048224.017

Britt, M. A., Rouet, J. F. & Durik, A. (2018). Representations and processes in multiple source use. In J. L. G. Braasch, I. Bråten, & M. T. McCrudden (Eds.), *Handbook of multiple source use* (pp. 17–33). New York, NY: Routledge.

Britt, M. A., & Rouet, J. F. (2020). Multiple document comprehension. In L. Zhang (Ed.), *The oxford encyclopedia of educational psychology*. Oxford: Oxford University Press.

Britt, M. A., Durik, A., & Rouet, J. F. (2022). Reading contexts, goals, and decisions: Text comprehension as a situated activity. *Discourse Processes*, *59*(5-6), 361–378. doi:10.1080/0163853X.2022.2068345

Britt, M. A., Perfetti, C. A., Sandak, R., & Rouet, J. F. (1999). Content integration and source separation in learning from multiple texts. In S. R. Goldman, A. C. Graesser & P. Van den Broek (Eds.), *Narrative comprehension, causality, and coherence: Essays in honor of Tom Trabasso* (pp. 209–233).

Mahwah, NJ: Lawrence Erlbaum Associates. doi:10.4324/9781410603135

Britt, M. A., Rouet, J. F., & Braasch, J. L. (2012). Documents as entities: Extending the situation model theory of comprehension. In M. A. Britt, J-F. Rouet & J. L. G. Braasch (Eds.) *Reading-from words to multiple texts* (pp. 174–193). New York, NY: Routledge. doi:10.4324/9780203131268

Burin, D. I., González, F. M., Martínez, M., & Marrujo, J. G. (2021). Expository multimedia comprehension in E-learning: Presentation format, verbal ability and working memory capacity. *Journal of Computer Assisted Learning*, *37*(3), 797–809. doi:10.1111/jcal.12524

Fernández Huerta, J. (1959). Medidas sencillas de lecturabilidad. *Consigna*, *214*, 29–32.

Foy, J. E., LoCasto, P. C., Briner, S. W., & Dyar, S. (2017). Would a madman have been so wise as this?" The effects of source credibility and message credibility on validation. *Memory & Cognition*, *45*(2), 281–295. doi:10.3758/s13421-016-0656-1

Gottschling, S., Kammerer, Y., & Gerjets, P. (2019). Readers' processing and use of source information as a function of its usefulness to explain conflicting scientific claims. *Discourse Processes*, *56*(5-6), 429–446. doi:10.1080/0163853X.2019.1610305

Gottschling, S., Kammerer, Y., & Gerjets, P. (2021). *Investigating Differences in Experts' and Laypersons' Evaluation of a Scientific Conflict with Eye-Tracking and Cued-Retrospective Thinking Aloud* (No. 6184). EasyChair. Retrieved from https://easychair.org/publications/preprint_open/3rzX.

Hargittai, E., Fullerton, L., Menchen-Trevino, E., & Thomas, K. Y. (2010). Trust online: Young adults' evaluation of web content. *International Journal of Communication*, *4*, 27. https://ijoc.org/index.php/ijoc/article/download/636/423.

Hassan, A., & Barber, S. J. (2021). The effects of repetition frequency on the illusory truth effect. *Cognitive Research: Principles and Implications*, *6*(1), 1–12. doi:10.1186/s41235-021-00301-5

Helsper, E. J., & Eynon, R. (2013). Distinct skill pathways to digital engagement. *European Journal of Communication*, *28*(6), 696–713. doi:10.1177/0267323113499113

Just, M. A., Carpenter, P. A., & Woolley, J. D. (1982). Paradigms and processes in reading comprehension. *Journal of Experimental Psychology. General*, *111*(2), 228–238. doi:10.1037/0096-3445.111.2.228

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*(5-6), 430–456. doi:10.1080/0163853X.2016.1169968

Khan, A. M., Khan, H. M., Huang, P., Warrian, K., & Gooi, P. (2022). Assessing the quality of online information on glaucoma procedures. *Canadian Journal of Ophthalmology. Journal Canadien D'ophtalmologie*, *57*(1), 23–28. doi:10.1016/j.jcjo.2021.02.010

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmerTest package: Tests in linear mixed effects models. *Journal of Statistical Software*, *82*(13), 1–26. doi:10.18637/jss.v082.i13

Lim, M. S., Molenaar, A., Brennan, L., Reid, M., & McCaffrey, T. (2022). Young adults' use of different social media platforms for health information: Insights from web-

based conversations. *Journal of Medical Internet Research*, *24*(1), e23656. doi:10.2196/23656

List, A., & Alexander, P. A. (2017). Cognitive affective engagement model of multiple source use. *Educational Psychologist*, *52*(3), 182–199. doi:10.1080/00461520.2017.1329014

List, A., Alexander, P. A., & Stephens, L. A. (2017). Trust but verify: Examining the association between students' sourcing behaviors and ratings of text trustworthiness. *Discourse Processes*, *54*(2), 83–104. doi:10.1080/0163853X.2016.1174654

List, A., Lee, H. Y., Du, H., Oaxaca, G. S. C., Lyu, B., Falcon, A. L., & Lin, C. J. (2022). Preservice teachers' recognition of source and content bias in educational application (app) reviews. *Computers in Human Behavior*, *134*, 107297.doi:10.1016/j.chb.2022.107297

Lo, S., & Andrews, S. (2015). To transform or not to transform: Using generalized linear mixed models to analyse reaction time data. *Frontiers in Psychology*, *6*, 1171. doi:10.3389/fpsyg.2015.01171

Macedo-Rouet, M., Potocki, A., Scharrer, L., Ros, C., Stadtler, M., Salmerón, L., & Rouet, J. F. (2019). How good is this page? Benefits and limits of prompting on adolescents' evaluation of web information quality. *Reading Research Quarterly*, *54*(3), 299–321. doi:10.1002/rrq.241

Mahmood, K. (2016). Do people overestimate their information literacy skills? A systematic review of empirical evidence on the Dunning-Kruger effect. *Communications in Information Literacy*, *10*(2), 199–213. doi:10.15760/comminfolit.2016.10.2.24

Marlon, J., Maibach, E., Ballew, M., Bergquist, P., Goldberg, M., Gustafson, A., … Leiserowitz, A. (2020, June 4) *Who do americans trust most for information about COVID-19?* Yale Program on Climate Change Communication. Retrieved October 12, 2022, from https://climatecommunication.yale.edu/publications/who-do-americans-trust-most-for-information-about-covid-19/.

Mitchell, A., & Liedke, J. (2021, August 24). About four-in-ten Americans say social media is an important way of following covid-19 vaccine news. Pew Research Center. Retrieved February 15, 2022, from https://www.pewresearch.org/fact-tank/2021/08/24/about-four-in-ten-americans-say-social-media-is-animportant-way-of-following-covid-19-vaccine-news/.

Naeem, S. B., Bhatti, R., & Khan, A. (2021). An exploration of how fake news is taking over social media and putting public health at risk. *Health Information and Libraries Journal*, *38*(2), 143–149. doi:10.1111/hir.12320

Ng, J. Y., Ahmed, S., & Zhang, C. J. (2021). Dietary and herbal supplements for weight loss: Assessing the quality of patient information online. *Nutrition Journal*, *20*(1), 72–13. doi:10.1186/s12937-021-00729-x

OECD. (2021). *21st-century readers: Developing literacy skills in a digital world*. Paris: OECD Publishing. doi:10.1787/a83d84cb-en

Pérez, A., Potocki, A., Stadtler, M., Macedo-Rouet, M., Paul, J., Salmerón, L., & Rouet, J. F. (2018). Fostering teenagers' assessment of information reliability: Effects

of a classroom intervention focused on critical source dimensions. *Learning and Instruction*, *58*, 53–64. doi:10.1016/j.learninstruc.2018.04.006

Perfetti, C. A., Rouet, J. F., & Britt, M. A. (1999). Toward a theory of documents representation. In H. Van Oostendorp & S. R. Goldman (Eds.), *The construction of mental representations during reading* (pp. 99–122). Mahwah, NJ: Erlbaum. R Core Team. (2021). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. Retrieved from https://www.R-project.org/.

Rapp, D. N., & Salovich, N. A. (2018). Can't we just disregard fake news? The consequences of exposure to inaccurate information. *Policy Insights from the Behavioral and Brain Sciences*, *5*(2), 232–239. doi:10.1177/2372732218785193

Richter, T., & Maier, J. (2018). The role of validation in multiple-document comprehension. In T. Richter & J. Maier (Eds.) *Handbook of multiple source use* (pp. 151–167). New York, NY: Routledge.

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*(3), 478–493. doi:10.1037/0022-0663.88.3.478
Rouet, J. F., Le Bigot, L., de Pereyra, G., & Britt, M. A. (2016). Whose story is this? Discrepancy triggers readers' attention to source information in short narratives. *Reading and Writing*, *29*(8), 1549–1570. doi:10.1007/s11145-016-9625-0

Rouet, J. F., Ros, C., Bordas, B., Sanchiz, M., Saux, G., Richter, T., & Britt, M. A. (2020). *When Does Source Information Help? Content Vs. Source-Based Validation as a Function of Readers' Prior Knowledge* (*No. 4493)* EasyChair. Retrieved from https://easychair.org/publications/preprint/2w2N.

Rouet, J. F., Skov, O., De Pereyra, G., Ros, C., Le Bigot, L., & Vibert, N. (2018). Relevance versus big numbers: Students' criteria for selecting scholarly references online. *Journal of Experimental Psychology: Applied*, *24*(4), 476–489. https://psycnet.apa.org/doi/10.1037/xap0000194.

Salmerón, L., Delgado, P., & Mason, L. (2020). Using eye-movement modelling examples to improve critical reading of multiple webpages on a conflicting topic. *Journal of Computer Assisted Learning*, *36*(6), 1038–1051. doi:10.1111/jcal.12458

Salmerón, L., Strømsø, H. I., Kammerer, Y., Stadtler, M., & van den Broek, P. (2018). 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.
Saux, G., Britt, M. A., Vibert, N., & Rouet, J. F. (2021a). Building mental models from multiple texts: How readers construct coherence from inconsistent sources. *Language and Linguistics Compass*, *15*(3), e12409. doi:10.1111/lnc3.12409

Saux, G., Vibert, N., Dampuré, J., Burin, D. I., Britt, M. A., & Rouet, J. F. (2021b). From simple agents to information sources: Readers' differential processing of story characters as a function of story consistency. *Acta Psychologica*, *212*, 103191. doi:10.1016/j.actpsy.2020.103191

Sperber, D., Clément, F., Heintz, C., Mascaro, O., Mercier, H., Origgi, G., & Wilson, D. (2010). Epistemic vigilance. *Mind & Language*, *25*(4), 359–393. doi:10.1111/j.1468-0017.2010.01394.x

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: 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). Cognitive Science Society.

Strømsø, H. I., & Bråten, I. (2014). Students' sourcing while reading and writing from multiple web documents. *Nordic Journal of Digital Literacy*, *9*(02), 92–111. doi:10.18261/ISSN1891-943X-2014-02-02

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*(3), 192–204. doi:10.1016/j.learninstruc.2009.02.001

Strømsø, H. I., Bråten, I., Britt, M. A., & Ferguson, L. E. (2013). Spontaneous sourcing among students reading multiple documents. *Cognition and Instruction*, *31*(2), 176–203. doi:10.1080/07370008.2013.769994

Van Boekel, M., Lassonde, K. A., O'Brien, E. J., & Kendeou, P. (2017). Source credibility and the processing of refutation texts. *Memory & Cognition*, *45*(1), 168–181. doi:10.3758/s13421-016-0649-0

Vidal-Abarca, E., Reyes, H., Gilabert, R., Calpe, J., Soria, E., & Graesser, A. C. (2002). ETAT: Expository text analysis tool. *Behavior Research Methods,Instruments, & Computers : A Journal of the Psychonomic Society, Inc*, *34*(1), 93–107. doi:10.3758/BF03195428

Wineburg, S., & McGrew, S. (2019). Lateral reading and the nature of expertise: Reading less and learning more when evaluating digital information. *Teachers College Record: The Voice of Scholarship in Education*, *121*(11), 1–40. doi:10.1177/016146811912101102