



Research Article

Measuring what matters: Investigating what new types of assessments reveal about students' online source evaluations

A growing number of educational interventions have shown that students can learn the strategies fact checkers use to efficiently evaluate online information. Measuring the effectiveness of these interventions has required new approaches to assessment because extant measures reveal too little about the processes students use to evaluate live internet sources. In this paper, we analyze two types of assessments developed to meet the need for new measures. We describe what these assessments reveal about student thinking and how they provide practitioners, policymakers, and researchers options for measuring participants' evaluative strategies.

Authors: Joel Breakstone (1), Sarah McGrew (2), Mark Smith (1)

Affiliations: (1) Graduate School of Education, Stanford University, USA, (2) College of Education, University of Maryland, USA

How to cite: Breakstone, J., McGrew, S., & Smith, M. (2024). Measuring what matters: Investigating what new types of assessments reveal about students' online source evaluations. *Harvard Kennedy School (HKS) Misinformation Review*, 5(1).

Received: April 26th, 2023. Accepted: January 11th, 2024. Published: February 12th, 2024.

Research question

- What do different types of assessments reveal about students' ability to evaluate online information?

Essay summary

- Educational interventions in a variety of contexts have shown that students can learn the strategies professional fact checkers use to evaluate the credibility of online sources. Researchers conducting these interventions have developed new kinds of assessments—instruments that measure participants' knowledge, behaviors, or cognitive processes—to test the effects of their interventions.
- These new kinds of assessments are necessary because assessments commonly used to measure outcomes in misinformation research offer limited insights into participants' reasoning. Extant measures do not reveal whether students deploy effective evaluation strategies and do not tap whether students engage in common evaluative mistakes like judging surface-level features (e.g., a source's top-level domain or appearance).
- In this study, we investigated what new assessments revealed about how students evaluated online sources. Rather than replicate the findings of prior intervention studies, this study focused

¹ A publication of the Shorenstein Center on Media, Politics and Public Policy at Harvard University, John F. Kennedy School of Government.

on understanding what these assessments revealed about students' reasoning as they evaluated online information.

- The findings showed that the assessments were effective in revealing patterns in students' reasoning as they evaluated websites. Responses pointed to common challenges students encountered when evaluating online content and showed evidence of students' effective evaluation strategies.
- This study highlights possibilities for types of assessments that can be both readily implemented and provide insight into students' thinking. Policymakers could use similar tasks to assess program effectiveness; researchers could utilize them as outcome measures in studies; and teachers could employ them for formative assessment of student learning.

Implications

Increasing individuals' resiliency against harmful online content has been a focus of substantial recent research. Interventions have ranged in scope from brief YouTube video tutorials (Roozenbeek et al., 2022) to game-based simulations (Basol et al., 2021; Maertens et al., 2021; Roozenbeek & van der Linden, 2019, 2020) to teacher-led, school-based interventions (Pavlounis et al., 2022; Wineburg et al., 2022). In contrast, tools to assess individuals' ability to evaluate online sources remain underdeveloped. As Camargo and Simon (2022) noted about the field, there is a need to "become more demanding when it comes to the theory and operationalization of what is being measured" (p. 5). There is a particular need for assessments that provide insights into the types of strategies individuals use to evaluate online content.

A growing number of interventions have attempted to teach students to evaluate online information as fact checkers do. Research with professional fact checkers demonstrated that the most efficient, effective way to both identify misinformation and find reliable information was to investigate the source (Wineburg & McGrew, 2019). To do so, fact checkers engaged in *lateral reading*: they left unfamiliar sites, opened new browser tabs, and searched for information about the original site. When fact checkers read laterally, they relied on a small set of strategies to efficiently come to better conclusions, such as using Wikipedia to quickly learn about an unfamiliar source. Based on these findings, researchers in countries including the United States, Canada, and Sweden have developed interventions to teach people to employ strategies like lateral reading (e.g., Axelsson, 2021; Breakstone, Smith, Connors, et al., 2021; Brodsky et al., 2021; Kohnen et al., 2020; McGrew, 2020; McGrew & Breakstone, 2023; McGrew et al., 2019; Pavlounis et al., 2022; Wineburg et al., 2022).

In contrast to fact checkers' effective search strategies, a variety of studies have shown clear patterns in the types of mistakes people commonly make when they evaluate online sources (Bakke, 2020; Barzilai & Zohar, 2012; Breakstone, Smith, Wineburg, et al., 2021; Dissen et al., 2021; Gasser et al., 2012; Hargittai et al., 2010; Lurie & Mustafaraj, 2018; Martzoukou et al., 2020; McGrew, 2020; McGrew et al., 2018). Instead of reading laterally, people rely on *weak heuristics*, or evaluative approaches in which judgments are based on features of sources not directly tied to credibility. People base their evaluations primarily on surface-level features, like the appearance of a source or its content (Breakstone, Smith, Wineburg, et al., 2021; Wineburg et al., 2022). In fact, academics and college students who were included in the study with fact checkers (Wineburg & McGrew, 2019) often came to incorrect conclusions about sources because they based their evaluations on weak heuristics.

Extant assessments are not ideally suited for measuring evaluative strategies like lateral reading that require evaluators to go deeper than examining the content or appearance of a site. Some of the most influential studies in the field have assessed participants' skills by asking them to evaluate sources that are static (not connected to the internet) or hypothetical (created by researchers to simulate real sources).

For example, researchers have asked participants judge news headlines (Moore & Hancock, 2022; Pennycook, Binnendyk, et al., 2021; Pennycook, Epstein, et al., 2021), evaluate fictional sources (Roozenbeek & van der Linden, 2019), respond to multiple-choice questions (Roozenbeek et al., 2022), evaluate static versions of online sources (Cook et al., 2017), and evaluate “web-like” documents (Macedo-Rouet et al., 2019). Assessments like these focus participants’ attention on the online content itself. Consider the pre- and post-measures from a gamified intervention (Basol et al., 2020) in which participants learned about misinformation strategies (e.g., emotional content or discrediting opponents) and then rated the credibility of screenshots of fictional tweets designed to embody these misinformation strategies. Because they could only access a screenshot of a given tweet, participants’ ratings were based on the content of the tweet alone. An assessment like this does not provide evidence about a participant’s ability to leave the post and research it online. In fact, reading laterally may lead to markedly different conclusions. One of the study’s fictional tweets, an example of “polarization,” read: “Worldwide rise of left-wing extremist groups damaging world economy: UN report.” A similar headline actually appeared in *Foreign Policy*: “Far-right extremism is a global problem” (Ashby, 2021). Rather than rejecting this well-regarded source based on a headline, as the gamified intervention study wanted participants to do with a similar headline, a quick search online would confirm that *Foreign Policy* is a credible source of information.

Extant assessments also provide researchers with little information about the processes participants use to evaluate sources. Researchers in the field have relied heavily on Likert-type scales that ask participants to rate whether a source is accurate or credible (e.g., Pennycook, Epstein, et al., 2021) or to self-report behaviors, such as how likely they would be to share or like a social media post (e.g., Roozenbeek & van der Linden, 2020). Interventions that investigate changes in students’ evaluative strategies need assessments that require more than a numerical rating as evidence of student thinking.

Measuring the efficacy of interventions designed to teach fact checkers’ strategies required different types of assessments. To address this need, we developed new tasks. We designed constructed-response items that sought to measure participants’ ability to evaluate actual online content with a live internet connection and to identify when participants made common mistakes. One constructed-response item asked students to evaluate the credibility of an unfamiliar website (see Figure 1) that contained many features that students mistakenly associate with credibility. In one version of the task, students were asked to imagine that they were doing research about global warming and encountered the website of an organization called Friends of Science (friendsofscience.org). The question directed students to use any online resources, including new searches, to determine whether the website was a credible source of information. Although the site has surface-level features that students often incorrectly use as markers of credibility (e.g., dot-org URL, references to scientific studies, and staff with advanced degrees), students needed to leave the site and read laterally to identify its backers and purpose. Such a search reveals that the website, which argues that humans are not responsible for climate change, receives funding from the fossil fuel industry.

You are researching global warming and come across this website:
<https://friendsofscience.org>. Please decide if this website is a trustworthy source of information on global warming. You can open a new tab and do an internet search if that helps.

Is this website a trustworthy source for learning about global warming?

Explain your answer, citing evidence from the webpages you used. Be sure to provide the URLs to the webpages you cite.

You are researching global warming and come across this website:
<http://www.co2science.org>. Please decide if this website is a trustworthy source of information on global warming. You can open a new tab and do an internet search if that helps.

Is this website a trustworthy source for learning about global warming?

Explain your answer, citing evidence from the webpages you used. Be sure to provide the URLs to the webpages you cite.

Figure 1. Constructed-response questions.

Constructed-response items can provide rich information about participant thinking, but they are time-consuming to implement and labor-intensive to score. To create more efficient assessment options, we converted constructed-response items that we had piloted extensively in prior research (cf. Breakstone, Smith, Wineburg, et al., 2021; McGrew et al., 2018) into multiple-choice questions that could be scored much more quickly but still yielded evidence about participants' evaluation strategies. For example, we converted a constructed-response item that asked students to consider two online sources and decide which would be a better place to start research on a given topic. The task was designed to assess whether students relied on key misconceptions about online information, including that Wikipedia is never a good source and sites with .edu top-level domains are always strong sources. In one version of the task (see Figure 2), students were told to imagine that they were doing research on gun control and came across two different websites: (1) A Wikipedia entry with more than two hundred linked citations; and (2) a personal webpage from Duke University where someone had posted a National Rifle Association broadside detailing gun control "myths." Students were asked which source would be a better starting place for research and were then presented with four answer choices. The multiple-choice version we created included the same question stem and internet sources as the constructed-response item, but

rather than generate a written response, participants were asked to identify which line of reasoning was correct (see Figure 2). Only one answer choice reflected correct reasoning about both sources; the three distractors featured the incorrect lines of reasoning about both sources that appeared most frequently in students' responses to the constructed-response version of the item.

Imagine you are doing research on gun control and you find the following webpages. Because webpages can change frequently, these links will take you to archived versions of the articles. Please do not consider that in your evaluations.

Webpage A: <https://people.duke.edu/~gsmith/articles/myths.htm>

Webpage B: https://en.wikipedia.org/wiki/Gun_politics_in_the_United_States

Which of these two webpages do you think is a better place to start your research?

- Webpage A is better because it is from Duke University, and Webpage B is worse because anyone can edit a Wikipedia page
- Webpage A is better because it is a .edu site, and Webpage B is worse because Wikipedia is never a good source to use
- Webpage B is better because it has more information than Webpage A
- Webpage B is better because it links to reliable references, and Webpage A is worse because its content is from an unreliable organization

Imagine you are doing research on criminal justice reform and you find the following webpages.

Webpage A: <https://sites.psu.edu/aspsy/2019/10/20/criminal-justice-reform/>

Webpage B: https://en.wikipedia.org/wiki/Criminal_justice_reform_United_States

Which of these two webpages do you think is a better place to start your research?

- Webpage A is better because it is from Penn State University, and Webpage B is worse because anyone can edit a Wikipedia page
- Webpage A is better because it is a .edu site, and Webpage B is worse because Wikipedia is never a good source to use
- Webpage B is better because it has more information than Webpage A
- Webpage B is better because it links to reliable references, and Webpage A is worse because its content is from an unreliable source

Figure 2. Multiple-choice questions.

As part of an intervention in high school classrooms, students completed these tasks before and after a series of lessons focused on teaching fact checker strategies like lateral reading. We analyzed how the tasks performed in the context of an intervention so we could understand not only what they revealed about student thinking but whether they were sensitive to student learning of effective evaluation strategies. Our purpose was to investigate what the assessments revealed in this context, not to gauge the effectiveness of the intervention (as we have done in prior studies of similar interventions; see McGrew & Breakstone, 2023; Wineburg et al., 2022). Analysis of student responses revealed that these assessments provided evidence of the reasoning students used to evaluate online information. The items

allowed us to gauge student proficiency in evaluating online sources and to identify the most common mistakes students made, such as judging sites by their appearance or top-level domain. Both item formats also showed sensitivity to changes in student reasoning after an intervention that taught lateral reading. That is, both item formats detected shifts in the reasoning processes that the intervention aimed to teach, which suggests the items were effective in measuring what students had learned.

These two types of assessments are most likely optimal for different applications. The short constructed-response item format is best suited for applications in which researchers and educators need to see the kinds of reasoning students generate and have the time and resources needed to read and classify responses. For example, classroom teachers might use constructed-response items as an informal formative assessment to glean insights about how their students evaluate online sources, using similar assessments over time to track changes in students' evaluations as they learn strategies like lateral reading. The multiple-choice format can be administered and scored much more quickly, so it offers advantages for situations in which educators and researchers need to gauge student or participant thinking quickly and inexpensively. For example, the multiple-choice format could be more effective than constructed-response items for an assessment that monitors student growth for thousands of students across an entire school district or state.

Findings

Finding 1: Student responses to a constructed-response item revealed students' reasoning as they evaluated a website.

A constructed-response item gauged whether students could successfully read laterally to investigate the source of a website. This task was designed to provide insight into students' ability to engage in an evaluative strategy that requires looking past surface features of a site and actively searching for information about the site's source. Students were asked to decide whether a website was a reliable source of information on climate change (see description above and Figure 1); on both forms of the item, the most efficient way to discover that the website was backed by the fossil fuel industry was to read laterally.

Students' pretest responses revealed how they approached evaluating the site before the intervention. On the pretest, eight responses out of 109 (7%) showed evidence of attempts to read laterally to investigate the source of the website. Meanwhile, 93% of responses ($n = 101$) did not show evidence of leaving the site to investigate its source. Instead, most of these responses revealed reliance on weak heuristics, or drew conclusions about credibility based on fallible surface features of the website being evaluated. As Table 1 shows, responses tended to rely on at least one of four weak heuristics: judging the site based on the About page's description of the source, the appearance of the site, its content, or the site's top-level domain. Additionally, the open-ended nature of this item's format allowed us to see that most pretest responses—61 responses (56%)—referenced more than one weak heuristic.

Table 1. Weak heuristics in pretest responses.

Judge reliability based on:	Pretest Responses	Sample Response
Content: quantity or perceived quality of information contained within the site	55	"It is trustworthy because it provides deep, background information and what scientists are predicting will happen in the future. It also quotes some scientists and explains their different claims. http://www.co2science.org/data/acidification/background.php "
URL: features of the site's URL, especially the top-level domain	43	"Yes, this website is a trustworthy source to learn about global warming. This is because any website that has .org after it is usually a credible source."
Appearance: features of the site's appearance, including its site design and visual appeal	34	"I feel like that website is kinda not trustworthy because it was so unorganized and there was so much going on in the middle."
About Page: organization's About page description	29	"It says on the website, 'Friends of Science is a non-profit organization run by dedicated volunteers comprised mainly of active and retired earth and atmospheric scientists, engineers, and other professionals.' This means that all the information on this website are from professionals. https://friendsofscience.org/index.php?id=1 "

Notes: Two raters coded responses independently and interrater reliability was high across all four codes, Cohen's $\kappa = .912$ (95% CI, .871 to .953), $p < .001$.

Students' posttest responses shed light on the extent to which students engaged in the evaluation strategy on which the intervention focused—lateral reading. On the posttest, 67 responses (61%) showed evidence of attempts to read laterally to investigate the source of the website. Meanwhile, 42 responses (39%) did not show evidence of lateral reading. Within responses that showed evidence of attempts to read laterally, the item format allowed us to differentiate between lateral reading that was ultimately successful and lateral reading that was not successful. Many responses (see Table 2) suggested that students successfully located evidence about the websites from external sources and reasoned about how that evidence cast doubt on the website's credibility. However, some students attempted to read laterally but came to inaccurate conclusions. As Table 2 shows, some responses showed evidence of students incorrectly concluding the site was a credible source of information because engineers worked for the organization or because it received funding from large companies like Exxon Mobil. Other responses referenced sources like Wikipedia but did not show evidence of gleaning the information most relevant to the site's reliability from those sources.

Table 2. Attempts at lateral reading in posttest responses.

Approach	Posttest Responses	Sample Response
Successful Lateral Reading		
	38	“According to Wikipedia, they are heavily funded by the petroleum industry which influences them to spread false information on global warming.”
Unsuccessful Lateral Reading		
Experts: Source is credible due to perceived expertise of staff	12	“Yes, it is because it is made up of active and retired engineers. https://www.sourcewatch.org/index.php/Friends_of_Science ”
Incorrect conclusion: Locates website that contains relevant information about source credibility but does not identify that information	9	“I looked up the name of the website and found a link to Wikipedia. This told me that the site is reliable. https://en.wikipedia.org/wiki/Friends_of_Science ”
Funding: Website is credible because of the nature of its funding	3	“This is a reliable source because it is run by a larger company that is trusted by Exxon Mobil. According to Wikipedia, this is a non-profit source and it is funded by America’s biggest coal mining company. https://en.wikipedia.org/wiki/Center_for_the_Study_of_Carbon_Dioxide_and_Global_Change ”
Irrelevant: No information related to source credibility is located	2	“It is very biased and the people that run the blog don’t have much credit. https://www.linkedin.com/in/ken-gregory-04bb543 states that he is an IT professional, which is not what he claims on the website.”

Notes: Two raters independently coded whether lateral reading responses were “successful” or “unsuccessful,” and interrater agreement was 100%. The same two raters coded responses for the “unsuccessful” processes students used. Interrater reliability was high across all four codes, Cohen’s $\kappa = .842$ (95% CI, .720 to .964), $p < .001$.

The constructed-response assessment allowed us to track how students’ evaluation strategies changed from pretest to posttest, including whether students engaged in lateral reading or relied on weak heuristics when evaluating sources. For example, on the pretest, a student who relied on weak heuristics argued that the linked site was reliable: “The website design is one used commonly in cheaper projects. However, it’s a .org website, there are a lot of seemingly usable data, and it seems like the entire website focuses on their goal.” This student relied on several weak heuristics—the site’s appearance, content, and URL—to judge the site. At posttest, the same student showed evidence of reading laterally about the linked site and argued that it was not reliable. They explained, “They reject scientific evidence and instead propose that ‘the Sun is the main direct and indirect driver of climate change,’ not human activity (Wikipedia). They are also largely funded by the fossil fuel industry.”

Finding 2: Student responses to a multiple-choice item also allowed for tracking changes in students' evaluative approaches.

Multiple-choice items were also effective in revealing changes in student thinking from pre- to posttest. Because we based the answer choices on common student responses to previously piloted constructed-response versions of the task, this assessment format provided quick insight into students' evaluative approaches. Table 3 shows the pre/post results for a multiple-choice item that asked students to determine which of two sources was better for researching an important social issue (Form A: animal testing; Form B: gun control; see items in Figure 2). For both items, students had to choose between a Wikipedia entry with a robust list of references or a problematic source with a dot-edu top-level domain. The proportion of students who answered correctly improved from 6% to 59%, suggesting a significant shift in student beliefs about Wikipedia after the intervention.

The multiple-choice distractors (answers A, B, and C in Table 3) included the most common mistakes students made in response to constructed-response items in prior research (cf. Breakstone, Smith, Wineburg, et al., 2021; McGrew et al., 2018). Results for the multiple-choice items suggested that students were less likely to engage in these problematic strategies at posttest. Answer A, for example, featured two misconceptions: that Wikipedia does not control who edits its content and that university webpages are always reliable. The proportion of students who selected this answer choice declined from 52% to 34%. Answer choice B featured additional common misconceptions: that Wikipedia is never a useful source and that the top-level domain is a reliable indicator of credibility. The proportion of students who selected this item plummeted from 40% at pretest to only 1% at posttest, providing useful information about how students reason about online sources and whether the intervention dispelled common myths.

Table 3. Multiple-choice responses at pretest and posttest.

Question: Imagine you are doing research on [topic varies by question] and you find the following webpages. Which of these two webpages do you think is a better place to start your research?

Answer choices	Pretest (n = 115)	Posttest (n = 115)
<i>Answer A</i> – Webpage A is better because it is from a reputable university, and Webpage B is worse because anyone can edit a Wikipedia page (INCORRECT)	52%	34%
<i>Answer B</i> – Webpage A is better because it is a .edu site, and Webpage B is worse because Wikipedia is never a good source to use (INCORRECT)	40%	1%
<i>Answer C</i> – Webpage B is better because it has more information than Webpage A (INCORRECT)	2%	6%
<i>Answer D</i> – Webpage B is better because it links to reliable sources, and Webpage A is worse because its content is from an unreliable source (CORRECT)	6%	59%

Methods

In this study, we investigated what online source evaluation assessment items reveal about students' reasoning as they evaluated digital content. Data were drawn from a broader project in which ninth-grade biology and world geography teachers collaborated with the research team to teach a series of lessons on evaluating online information over the course of a school year. Our analysis focused on student responses to two items embedded in a pretest and a posttest administered to one of the participating biology teacher's students. We focused on responses from this teacher's students because the teacher was involved in the design of the lessons and taught all the lessons the team designed. The teacher taught five hour-long lessons on evaluating online information over the course of the 2021–22 school year. The lessons introduced students to the importance of asking, "Who is behind this information?" when they encountered online content from an unfamiliar source and to engage in lateral reading to investigate the source. For example, in the first lesson, which took place during an introductory biology unit about caffeine, students learned to read laterally by watching a screencast demonstrating lateral reading about an article regarding the benefits of caffeine from a website funded by beverage industry groups. Students then practiced reading laterally about another article about caffeine from a crowd-sourced website that lacks editorial review. The lessons also explicitly addressed common mistakes students might make when evaluating online information and introduced resources to which students could turn when reading laterally, including Wikipedia. (See Appendix for a description of the learning objectives and activities in each lesson.)

Assessment design and implementation

We designed two forms of an assessment—Form A and Form B—to assess student thinking before any of the lessons and after all five were taught. Both forms were composed of measures of students' ability to evaluate digital content that were piloted extensively (McGrew et al., 2018) and have been used as measures of student thinking in other intervention studies (e.g., Breakstone, Smith, Connors, et al., 2021; Wineburg et al., 2022). The items assessed students' ability to critically evaluate online sources and to read laterally to investigate a source's credibility.

The assessment included three multiple-choice items and three constructed-response items. Due to space constraints, we discuss two items—one constructed-response and one multiple-choice—here. Although they varied in the kind of response they required from students, both constructed-response and multiple-choice items engaged students in evaluating actual online content by linking to websites for students to evaluate. Each item on Form B was designed to be parallel to an item on Form A. Parallel items contained the same question stem (and, for the multiple-choice items, the same answer choices) but asked students to evaluate different examples of online content.

Students completed the pre- and posttests via a Qualtrics form during their biology class. The assessments featured live internet sources, and students were allowed to go online at any time while completing them. We used a randomizer in Qualtrics to assign forms to students at pretest, and students received the opposite form at posttest.

Analysis

Constructed-response items were analyzed and coded for the strategies students used in their answers. We developed an initial coding scheme based on prior experience with the items (e.g., McGrew, 2020) as well as existing research on how students evaluate online information (e.g., Breakstone, Smith, Wineburg, et al., 2021; McGrew et al., 2018). We added and refined codes to represent the full range of reasoning

we saw in responses to each item. Two raters then coded each response using the final coding scheme. Estimates of interrater reliability were high at both pretest and posttest administrations of the constructed-response items (see estimates in Table 1 and Table 2).

Multiple-choice items were scored and the percent of correct responses—as well as incorrect responses of different varieties—were tabulated. Finally, responses and codes for both assessment formats were analyzed to describe student reasoning revealed in the responses and to understand how, if at all, reasoning changed from pre- to posttest.

Limitations and future directions

Although this analysis suggested that both assessment formats were effective in revealing student thinking and changes in students' evaluative processes, it does not provide evidence about the equivalence of these assessment formats. Each of the assessments featured different stimulus materials and targeted different evaluation skills, which limited our ability to evaluate the effects of question format on item functioning. Examining the effectiveness of different assessment formats is a worthwhile avenue for continued inquiry. For example, researchers could compare how students respond to constructed-response and multiple-choice response options for the same question with the same stimuli.

Although there are a variety of ways these assessment formats might be used by educators, policymakers, and researchers, this study did not provide evidence of their efficacy in these varied real-world settings. Further research is needed to explore how these assessment tools are used in practice.

Bibliography

- Ashby, H. (2021, January 15). *Far-right extremism is a global problem*. Foreign Policy. <https://foreignpolicy.com/2021/01/15/far-right-extremism-global-problem-worldwide-solutions/>
- Axelsson, C. W., Guath, M., & Nygren, T. (2021). Learning how to separate fake from real news: Scalable digital tutorials promoting students' civic online reasoning. *Future Internet*, 13(3), 60–78. <https://doi.org/10.3390/fi13030060>
- Bakke, A. (2020). Everyday googling: Results of an observational study and applications for teaching algorithmic literacy. *Computers and Composition*, 57, 102577–102516. <https://doi.org/10.1016/j.compcom.2020.102577>
- Barzilai, S., & Zohar, A. (2012). Epistemic thinking in action: Evaluating and integrating online sources. *Cognition and Instruction*, 30(1), 39–85. <https://doi.org/10.1080/07370008.2011.636495>
- Basol, M., Roozenbeek, J., Berriche, M., Uenal, F., McClanahan, W. P., & van der Linden, S. (2021). Towards psychological herd immunity: Cross-cultural evidence for two prebunking interventions against COVID-19 misinformation. *Big Data & Society*, 8(1). <https://doi.org/10.1177/20539517211013868>
- Breakstone, J., Smith, M., Connors, P., Ortega, T., Kerr, D., & Wineburg, S. (2021). Lateral reading: College students learn to critically evaluate internet sources in an online course. *Harvard Kennedy School (HKS) Misinformation Review*, 2(1). <https://doi.org/10.37016/mr-2020-56>
- Breakstone, J., Smith, M., Wineburg, S., Rapaport, A., Carle, J., Garland, M., & Saavedra, A. (2021). Students' civic online reasoning: A national portrait. *Educational Researcher*, 50(8), 505–515. <https://doi.org/10.3102/0013189X211017495>

- Brodsky, J. E., Brooks, P. J., Scimeca, D., Todorova, R., Galati, P., Batson, M., Grosso, R., Matthews, M., Miller, V., & Caulfield, M. (2021). Improving college students' fact-checking strategies through lateral reading instruction in a general education civics course. *Cognitive Research: Principles and Implications*, 6(23). <https://doi.org/10.1186/s41235-021-00291-4>
- Camargo, C. Q., & Simon, F. M. (2022). Mis- and disinformation studies are too big to fail: Six suggestions for the field's future. *Harvard Kennedy School (HKS) Misinformation Review*, 3(5). <https://doi.org/10.37016/mr-2020-106>
- Cook, J., Lewandowsky, S., & Ecker, U. K. H. (2017). Neutralizing misinformation through inoculation: Exposing misleading argumentation techniques reduces their influence. *PLOS ONE*, 12(5), e0175799. <https://doi.org/10.1371/journal.pone.0175799>
- Dissen, A., Qadiri, Q., & Middleton, C. J. (2021). I read it online: Understanding how undergraduate students assess the accuracy of online sources of health information. *American Journal of Lifestyle Medicine*, 16(5), 641–654. <https://doi.org/10.1177/1559827621990574>
- Gasser, U., Cortesi, S. C., Malik, M. M., & Lee, A. (2012). *Youth and digital media: From credibility to information quality*. Berkman Center Research Publication No. 2012-1. <https://doi.org/10.2139/ssrn.2005272>
- Hargittai, E., Fullerton, L., Menchen-Trevino, E., & Thomas, K. Y. (2010). Trust online: Young adults' evaluation of web content. *International Journal of Communication*, 4, 468–494. <https://ijoc.org/index.php/ijoc/article/view/636/423>
- Kohnen, A. M., Mertens, G. E., & Boehm, S. M. (2020). Can middle schoolers learn to read the web like experts? Possibilities and limits of a strategy-based intervention. *Journal of Media Literacy Education*, 12(2), 64–79. <https://doi.org/10.23860/JMLE-2020-12-2-6>
- Lurie, E., & Mustafaraj, E. (2018). Investigating the effects of Google's search engine result page in evaluating the credibility of online news sources. In *WebSci18: Proceedings of the 10th ACM conference on web science* (pp. 107–116). Association for Computing Machinery. <https://doi.org/10.1145/3201064.3201095>
- Macedo-Rouet, M., Potocki, A., Scharrer, L., Ros, C., Stadtler, M., Salmerón, L., & Rouet, J. (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. <https://doi.org/10.1002/rrq.241>
- Maertens, R., Roozenbeek, J., Basol, M., & van der Linden, S. (2021). Long-term effectiveness of inoculation against misinformation: Three longitudinal experiments. *Journal of Experimental Psychology: Applied*, 27(1), 1–16. <https://doi.org/10.1037/xap0000315>
- Martoukou, K., Fulton, C., Kostagiolas, P., & Lavranos, C. (2020). A study of higher education students' self-perceived digital competences for learning and everyday life online participation. *Journal of Documentation*, 76(6), 1413–1458. <https://doi.org/10.1108/JD-03-2020-0041>
- McGrew, S. (2020). Learning to evaluate: An intervention in civic online reasoning. *Computers and Education*, 145, 1–13. <https://doi.org/10.1016/j.compedu.2019.103711>
- McGrew, S., & Breakstone, J. (2023). Civic online reasoning across the curriculum: Developing and testing the efficacy of digital literacy lessons. *AERA Open*, 9. <https://doi.org/10.1177/23328584231176451>
- McGrew, S., Breakstone, J., Ortega, T., Smith, M., & Wineburg, S. (2018). Can students evaluate online sources? Learning from assessments of civic online reasoning. *Theory & Research in Social Education*, 46(2), 165–193. <https://doi.org/10.1080/00933104.2017.1416320>
- McGrew, S., Smith, M., Breakstone, J., Ortega, T., & Wineburg, S. (2019). Improving students' web savvy: An intervention study. *British Journal of Educational Psychology*, 89(3), 485–500. <https://doi.org/10.1111/bjep.12279>

- Moore, R. C., & Hancock, J. T. (2022). A digital media literacy intervention for older adults improves resilience to fake news. *Scientific Reports*, 12, 6008. <https://doi.org/10.1038/s41598-022-08437-0>
- Pavlounis, D., Johnston, J., Brodsky, J., & Brooks, P. (2021). *The digital media literacy gap: How to build widespread resilience to false and mis-leading information using evidence-based classroom tools*. CIVIX Canada. <https://ctrl-f.ca/en/wp-content/uploads/2022/01/The-Digital-Media-Literacy-Gap.pdf>
- Pennycook, G., Binnendyk, J., Newton, C., & Rand, D. (2021). A practical guide to doing behavioral research on fake news and misinformation. *Collabra: Psychology*, 7(1). <https://doi.org/10.1525/collabra.25293>
- Pennycook, G., Epstein, Z., Mosleh, M., Arechar, A. A., Eckles, D., & Rand, D. G. (2021). Shifting attention to accuracy can reduce misinformation online. *Nature*, 592, 590–595. <https://doi.org/10.1038/s41586-021-03344-2>
- Roozenbeek, J., & van der Linden, S. (2019). Fake news game confers psychological resistance against online misinformation. *Palgrave Communications*, 5(65). <https://doi.org/10.1057/s41599-019-0279-9>
- Roozenbeek, J., & van der Linden, S. (2020). Breaking Harmony Square: A game that “inoculates” against political misinformation. *Harvard Kennedy School (HKS) Misinformation Review*, 1(8). <https://doi.org/10.37016/mr-2020-47>
- Roozenbeek, J., van der Linden, S., Goldberg, B., Rathje, S., & Lewandowsky, S. (2022). *Science Advances*, 8(34). <http://doi.org/10.1126/sciadv.abo6254>
- Wineburg, S., Breakstone, J., McGrew, S., Smith, M., & Ortega, T. (2022). Lateral reading on the open Internet: A district-wide field study in high school government classes. *Journal of Educational Psychology*, 114(5), 893–909. <https://doi.org/10.1037/edu0000740>
- Wineburg, S., & McGrew, S. (2019). Lateral reading and the nature of expertise: Reading less and learning more when evaluating digital information. *Teachers College Record*, 121(11), 1–40. <https://doi.org/10.1177/016146811912101102>

Acknowledgements

The authors thank the educators who collaborated on this project.

Funding

This research was generously supported by funding from the Brinson Foundation, the Robert R. McCormick Foundation, and the Spencer Foundation, Sam Wineburg, Principal Investigator; however, the authors alone bear responsibility for the contents of this article.

Competing interests

The authors declare no competing interests.

Ethics

The research protocol employed was approved by the Stanford University Institutional Review Board.

Copyright

This is an open access article distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided that the original author and source are properly credited.

Data availability

All materials needed to replicate this study are available via the Harvard Dataverse:

<https://doi.org/10.7910/DVN/QLFCCQ>

Appendix: Intervention lessons and objectives

Table A1. Intervention lessons and objectives.

Topic	Description
Caffeine: Intro to lateral reading	As part of an introductory biology unit about caffeine, students learned to read laterally by watching a screencast demonstrating lateral reading about an article regarding the benefits of caffeine from a website funded by beverage industry groups. Students then practiced reading laterally about another article about caffeine from a crowd-sourced website that lacks editorial review.
Ecology: Lateral reading continued	Students practiced the skill of lateral reading by evaluating sites of varying quality during an ecology unit project that asked them to find information about a particular animal.
Nutrition: Lateral reading on social media part 1	Students practiced reading laterally on social media. The whole class read laterally about a TikTok video from a registered dietician. Students then individually created screencasts of themselves reading laterally about a source they found on social media.
GMOs: Lateral reading on social media part 2	Students practiced reading laterally on social media. During a unit on genetically modified organisms, they evaluated a post from an Instagram account called “GMO answers.” By reading laterally, they learned that the website is run by agrochemical companies with a vested interest in public policy about GMOs.
Forensics: Authority of sources for lateral reading	Students were introduced to the skill of click restraint. They were shown a search engine results page (SERP) and asked to consider which result they would select first. They then read laterally to investigate the sources.