

Judging a text by its author — A meta-analysis of interventions to foster source credibility assessment[☆]

Marvin Fendt^{ID*}, Xenia Muth^{ID}, Peter Adriaan Edelsbrunner^{ID}

Faculty of Psychology and Educational Sciences, Ludwig Maximilian University Munich, Germany

ARTICLE INFO

Keywords:

Sourcing
Lateral reading
Source credibility
Misinformation
Meta-analysis

ABSTRACT

Critical evaluation of source credibility is essential in today's digital landscape but often requires explicit instruction. Our meta-analysis synthesizes findings from 64 controlled experimental studies to assess the effectiveness of four different intervention approaches (i.e., historical thinking, multiple document literacy, sourcing, and lateral reading) to foster source credibility assessment. Source credibility assessment interventions were overall effective ($g = 0.42, p < .001$), with lateral reading showing the largest effects. We found considerable heterogeneity (95% Prediction Interval $[-0.33, 1.17]$), indicating that expected effect sizes in a random study from among the population of studies reviewed range from small negative to large positive effects. Greater effects were observed with graduated participants compared to other educational backgrounds, as well as in university and school settings, regardless of age and gender composition. Additionally, interventions that used the open Internet demonstrated enhanced effectiveness. We conclude that (1) lateral reading is particularly suited to the digital information landscape, (2) repeated practice may enhance intervention effectiveness, and (3) ecological validity is highly important for intervention effectiveness.

1. Educational relevance statement

In an era where citizens have to acquire the skills to identify credible information in a seemingly endless stream of content, this meta-analysis provides critical insights into the effectiveness of educational interventions to foster source credibility assessment. We found small to moderate overall effects across different approaches (historical thinking, sourcing, multiple document comparison, lateral reading). Lateral reading, an approach specifically designed for online sources, appears particularly effective at equipping learners for the demands of the digital information landscape. Our results underscore the importance of integrating repeated practice, using ecologically valid learning settings, and digital technologies to improve source credibility assessment across educational levels.

2. Introduction

Testaments of dubious information are not new and date back at least to the Roman Empire (Pennycook & Rand, 2022). Throughout history, the ability to judge the credibility of texts based on authorship or content has been predominantly reserved for scholars and experts (Graves, 2016). In recent centuries, however, the advent of

mass printing and, more recently, the Internet has democratized knowledge and information access, enabling greater participation in global discourse (Hunter, 2023; Kenski & Stroud, 2006). While traditional dissemination of information typically involved institutionalized content gatekeeping prior to publication, for example by reviewers and fact-checkers at publishing houses, such mechanisms are absent for most online content (Singer, 2023). These ungated information environments shift the burden of evaluating information credibility from professionals to users who may be unaware of this circumstance (Flanagin & Metzger, 2008).

We define sources as authors, publishers, institutions, or platforms that produce or distribute content (Perfetti, Rouet, & Britt, 1999; Wineburg, 1991). This means that an author can be any kind of originating source with primary authorship responsibility for a published text, such as a blogger or a publishing house. This definition excludes private redistribution of information, such as citizens who share information in their private lives.

Meaningful strategies to identify credible sources do not come naturally, even to many highly educated individuals and regardless of age, particularly in online environments (e.g., Breakstone et al., 2021; Kirschner & van Merriënboer, 2013; Wineburg, Breakstone, McGrew, Smith, & Ortega, 2022). Especially on social media platforms,

[☆] This article is part of a Special issue entitled: 'Metaanalyses & Systematic Reviews' published in Learning and Individual Differences.

* Correspondence to: Department of Educational Psychology, Ludwig Maximilian University, Leopoldstr. 13, D-80802 München, Germany.

E-mail address: marvin.fendt@psy.lmu.de (M. Fendt).

users often process source information only superficially or ignore it completely (Hämäläinen, Kiili, Räikkönen, Lakkala, Ilomäki, Toom, & Marttunen, 2023). Therefore, skills to read, understand, and evaluate texts need to be carefully taught (Osborne & Pimentel, 2022).

Over the past decades, research across various fields, notably persuasion, communication, literacy, and education, has developed domain-specific approaches to help individuals navigate the complexities of source credibility (Brante & Strømsø, 2018). The process and skill set of meaningfully evaluating source credibility is commonly referred to as sourcing or source credibility assessment, which we define as evaluating the credibility of a text using and crosschecking available first- and second-hand information about the source and its meta-characteristics, such as who created the document, when, and for what purpose (similar to Bråten, Stadtler, & Salmerón, 2018). In contrast, content evaluation involves assessing the internal logic, evidence, and coherence of the claims within the text itself. While in practice, skilled readers often blend these processes (Barzilai, Thomm, & Shlomi-Elouo, 2020), the interventions we analyze are specifically those designed to teach the skills of sourcing. For example, a reader might start reading a well-written article out of curiosity before systematically checking its author's credibility (Marten, Altmann, Baumgarten-Kelm, & Stadtler, 2025). We focus on interventions that train source credibility assessment. Individuals lacking these skills risk being misled by misinformation, with far-reaching consequences for both the individual and society. Misinformation can lead to social fragmentation, polarization, and the erosion of trust in institutions, ultimately undermining democratic processes and collective decision-making (Ecker et al., 2022).

Researchers have designed interventions to foster source credibility assessment mainly based on the four approaches historical thinking, sourcing, multiple document literacy, and lateral reading (as reviewed by Brante & Strømsø, 2018; McGrew, 2024). Historical thinking (Wineburg, 1991) is one of the earliest approaches to evaluating source credibility. Historians have long emphasized the importance of understanding the context, perspective, and motives behind a source instead of face values to judge its reliability. Other strategies have built on that foundation. Sourcing (Bråten, Stadtler, & Salmerón, 2018) involves asking critical questions about the author, the author's expertise, potential biases, and the context in which the information was produced. Multiple document literacy has expanded on these skills to navigating, interpreting, and synthesizing the content of various texts, combining source with content knowledge (Rouet & Britt, 2011). More recently, lateral reading (Wineburg & McGrew, 2019) has emerged as a strategy tailored for assessing the credibility of online information. Developed by observing professional fact-checkers, lateral reading involves leaving the site of the original text to cross-check the source against information from other websites, rather than focusing on the text content.

Individual differences, including prior knowledge, beliefs, motivation, and cognitive abilities, may predict individuals' source credibility assessment skills (as reviewed for multiple document literacy by Anmarkrud, Bråten, Florit, & Mason, 2022). For example, readers with more prior knowledge of content — whether from formal education or life experience — are better equipped to critically engage with claims, making it easier to identify biases or false information (List, Grossnickle, & Alexander, 2016).

In the present study, we are interested in reviewing the effectiveness of source credibility assessment interventions following the different approaches described. We conduct a meta-analysis to evaluate the overall effectiveness of these interventions in fostering source credibility assessment skills. By synthesizing findings across studies, this research seeks to uncover how source credibility assessment skills can be effectively taught to support individuals in navigating the challenges of today's information landscape.

3. Conceptualizations of source credibility assessment in different research fields

In the following, with a “concept” in the context of source credibility assessment research, we mean a theoretical concept (e.g., the outline of lateral reading by Wineburg & McGrew, 2019). We use the term “intervention” for its practical implementation (e.g., Fendt, Nistor, Scheibenzuber, & Artmann, 2023; Wineburg et al., 2022), “skills” for the (internal) skills taught and “strategies” for the actual practical actions to assess source credibility. That means, built on a concept, an intervention attempts to teach people the skills they can apply as strategies.

Initial research on source credibility assessment mostly stems from persuasion, information and communication, as well as literacy and discourse research (as reviewed by Brante & Strømsø, 2018). However, since fostering source credibility assessment skills often involves expert guidance, for example, in training sessions, there is a growing body of research on source credibility assessment that is grounded in pedagogy (e.g., Barzilai & Ka'adan, 2017; Wineburg et al., 2022).

Persuasion research distinguishes between source trustworthiness (i.e., honesty and reliability) and expertise (i.e., perceived knowledge or competence), which affect persuasion episodes (Bråten, Stadtler, & Salmerón, 2018; Pornpitakpan, 2004). Other models have explored more dimensions like perceived goodwill or source attractiveness, which encompasses features like similarity, familiarity, and likability (Eisend, 2006; McCroskey, 1966; Ohanian, 1990). Persuasion research models describe how individuals process persuasive messages and make decisions. Two influential models, the Heuristic-Systematic Model (Chaiken, 1987) and the Elaboration Likelihood Model (Petty & Cacioppo, 1986), outline two routes that individuals can take in persuasion episodes and often share a common idea: One route of active, motivated investigation and another of passive, unmotivated reception. When individuals are highly motivated and have the cognitive ability to process the message, they are more likely to scrutinize the source and other message characteristics (Petty & Cacioppo, 1986; Petty & Wegener, 1999). On this deliberate route, individuals analytically process and deeply engage with the content (Chen & Chaiken, 1999). If people are not motivated or distracted, they choose a fast heuristic route based on cognitive shortcuts and low-effort judgment of source characteristics. These peripheral, superficial cues include the perceived credibility, likability, authority or attractiveness of the source, rather than the content of the message or meaningful but more effortful features.

Literacy and discourse research focuses on the interplay between understanding source features and content. Studies indicated that knowledge about the source of a text, mainly the genre, the intention of the author, and the publishing year, influences the way students approach its content (Zwaan, 1994). Readers can attribute conflicting perspectives to different sources, which helps them actively evaluate the trustworthiness of each source and resolve the conflict to form a coherent understanding of the issue (Braasch, McCabe, & Daniel, 2016; Britt, Richter, & Rouet, 2014; Stadtler & Bromme, 2014). While there are several unique models in this research strand, the documents model framework (Britt, Perfetti, Sandak, & Rouet, 1999; Perfetti et al., 1999) is among the most influential, conceptualizing two levels of cognitive representation for understanding complex situations such as conflicting sources: the integrated mental model and the intertext model. The integrated mental model reflects connections across the semantic content of multiple texts, identifying key ideas, areas of agreement, and points of disagreement. The intertext model involves cognitively representing the source features of each text, such as document authorship, publication date, and genre, as “document nodes”. The source features are then linked to specific content and to each other to reflect relationships between the sources, such as agreement or conflict. Additionally, not only the main sources (i.e., individual document sources) but also the

main references of those documents should be assessed (Britt & Rouet, 2012; Strømsø, Bråten, Britt, & Ferguson, 2013).

Information and communication research has paid particular attention to readers' perceptions of source credibility. In line with behavioral models of human motivation, achievement, and behavior such as the expectancy-value theory (Wigfield & Eccles, 2000), readers pay more attention if they expect the personal impact of the content to be greater. With the vast amount of information available online, individuals are often faced with information overload, which can make source credibility assessment and finding important information more difficult (Kozyreva, Wineburg, Lewandowsky, & Hertwig, 2023). Therefore, similar to the Elaboration Likelihood Model (Petty & Cacioppo, 1986), only few would even undertake the effort to check the source, unless they are motivated or incentivized to do so (Metzger, 2007; Pennycook & Rand, 2019).

Metzger, Flanagin, and Medders (2010) identified five traditional criteria for evaluating online information: Accuracy (i.e., the degree to which a source is error-free), authority (i.e., the expertise of the author), objectivity (i.e., the degree to which it is fact or opinion), currency (i.e., how up-to-date the information is), and scope (i.e., the comprehensiveness of the source's information). In addition to traditional criteria, Metzger identified social and heuristic cues to credibility: Users tend to trust more reputable (i.e., familiar names and sources), endorsed (i.e., number of likes on social media), and self-confirmatory (i.e., consistent with a person's prior beliefs) sources, while expectancy violating and persuasive (e.g., advertising) sources are deemed less credible.

Among other relevant models in this line of research, the main model of Sundar (2008) focuses on the structural features that trigger heuristics that influence users' credibility judgments: modality of how the information is presented (e.g., audio, text, video), agency of the information source, interactivity or the degree to which users can serve as both source and receiver of information, navigability or the ease of locating relevant information. For example, showing appreciation for content in likes would trigger an engagement heuristic in users and more liked content would then be perceived as more trustworthy.

Educational research has contributed insights on instructional components and ideal guidance to foster source credibility assessment skills. Recent studies (e.g., Fendt et al., 2023; Wineburg et al., 2022) found that scaffolding, which is an instructional support to help learners acquire new skills, provides a substantial benefit to learning source credibility assessment. Furthermore, educational research can support source credibility assessment through fostering metacognition, which encompasses being aware of one's own cognitive abilities, their limitations, and the learning process (Barzilai & Ka'adan, 2017).

4. Different approaches to source credibility assessment interventions

A range of interventions has been developed and evaluated to foster source credibility assessment. We trace them mainly to four common, partially overlapping research fields as well as models and instructional approaches derived from them: Sourcing (e.g., Hämäläinen, Kiili, Marttunen, Rääkkönen, González-Ibáñez, & Leppänen, 2020; Kammerer, Amann, & Gerjets, 2015; Macedo-Rouet, Braasch, Britt, & Rouet, 2013), historical reasoning (e.g., De La Paz, 2005; Nokes, 2014; Reisman, 2012), multiple document literacy (e.g., Delgado, Stang Lund, Salmerón, & Bråten, 2020; Griffin, Jaeger, Britt, & Wiley, 2024; Sonia et al., 2022), and the recently described lateral reading (e.g., Brodsky, 2022; Fendt et al., 2023; Moore & Hancock, 2022).

Although the different source credibility assessment approaches often show partial conceptual overlap, we aim to categorize distinct approaches to empirically compare their effectiveness. We use the following definitions: Historical thinking (Wineburg, 1991) is the process of analyzing source credibility, similar to historians, by source (sourcing), context (contextualization), and content (corroboration).

Sourcing (Bråten, Stadtler, & Salmerón, 2018) is the process of identifying who created a document, when, where, and for what purpose, in order to assess its reliability, perspective, and potential bias. Multiple document literacy (Anmarkrud et al., 2022) is the ability to understand, interpret, and synthesize information from a variety of documents or sources, which often requires weighing differing perspectives and evidence. Lateral reading (Wineburg & McGrew, 2019) is a structured approach similar to sourcing that aims to mostly disregard context and content to judge source credibility mainly based on third-party verdicts.

4.1. Historical thinking

One of the first advances to describe a structured source credibility assessment approach, historical thinking, was made by Wineburg (1991). Contrary to traditional rote learning of history, which often emphasized memorizing dates and events over critically engaging with sources, the approach revolves around teaching students to approach historical texts with a critical mindset. Wineburg found that expert historians often engaged critically with sources via sourcing, contextualization, and corroboration to assess the reliability of historical documents. Thus, historical thinking requires considering the context in which documents were created, the perspectives they represent, and the evidence they provide. These skills allow historians to construct a nuanced understanding of historical events by critically evaluating the origin, context, and consistency of information across multiple documents.

Wineburg (1991) described the three strategies sourcing, corroboration, and contextualization. Sourcing involves looking at the source information of a document to judge the credibility of its content (described in detail in the next section). Corroboration means comparing and connecting the information from different texts to find contradictions and similarities and make a more objective judgment (Nokes, Dole, & Hacker, 2007). Contextualization is defined as imagining the context in which the document was written (e.g., historical, political, geographical) to infer the lens through which the author viewed the events they described (De La Paz, Felton, Monte-Sano, Croninger, Jackson, Deogracias, & Hoffman, 2014).

Interventions designed to promote historical thinking have shown promising results. A curriculum by Reisman (2012) has been widely used in history education to foster critical engagement with historical documents. Burnett and Cuevas (2023) successfully taught historical reasoning strategies to elementary school students. VanSledright (2002) highlights that students who receive explicit instruction on historical thinking seem more able to assess the credibility of historical documents and construct coherent historical narratives.

4.2. Sourcing

Sourcing specifically refers to the practice of examining the meta-characteristics of a source, such as who created the document, when, and for what purpose (Bråten, Stadtler, & Salmerón, 2018). It builds on the critical analysis of historical thinking to encourage students to go beyond the content and consider author motivation and credibility. Sourcing involves attending to, representing, and evaluating a document's origin, purpose, and context with the aim of assessing its credibility. While the definition of sourcing does not strictly include content evaluation, sourcing intervention often include a form of content investigation (e.g., Marten et al., 2025).

There is a notable body of research that has been reviewed several times: Brante and Strømsø (2018) conclude in their review that sourcing interventions are particularly effective at the secondary and post-secondary levels (e.g., Britt & Aglinskas, 2002; Pérez, Potocki, Stadtler, Macedo-Rouet, Paul, Salmerón, & Rouet, 2018; Wiley et al., 2009), while the effectiveness of interventions at elementary schools might be limited due to students' development of reflective thinking (e.g., Kuiper, Volman, & Terwel, 2009). Other reviews emphasize

the role of motivation and personal relevance to engage in sourcing, with motivated students experiencing higher and sustained learning gains (Anmarkrud et al., 2022; Bråten, Brante, & Strømsø, 2018). Notably, some interventions have also been adapted for digital environments, like Delgado et al. (2020), where users must navigate conflicting sources and assess the credibility of multiple hypertext documents.

4.3. Multiple document literacy

Multiple document literacy is the ability to navigate, interpret, and synthesize content from various texts, attending not only to the sources but also to content comprehension (Rouet & Britt, 2011). It is grounded in the idea that individual documents are likely biased and require individuals to critically evaluate the differences in how various sources represent the same information. This idea makes the skill particularly important in contexts where information is dispersed across various sources (Anmarkrud, Bråten, & Strømsø, 2014). Like historical thinking and sourcing, multiple document literacy helps people become more discerning readers by encouraging them to evaluate not only the content of documents but also the contexts in which they were produced.

Several studies have supported the efficacy of multiple document literacy interventions (e.g., Barzilai, Mor-Hagani, Zohar, Shlomi-Elou, & Ben-Yishai, 2020; Bråten, Brandmo, Ferguson, & Strømsø, 2022). For example, Griffin et al. (2024) found that students trained in multiple document literacy better synthesized and integrated information from multiple documents, leading to better overall comprehension and critical thinking skills. Furthermore, addressing text content (Stadtler & Bromme, 2014) and providing scaffolding materials (Bråten, Anmarkrud, Brandmo, & Strømsø, 2014) can further boost the effectiveness of source credibility assessment interventions.

4.4. Lateral reading

Wineburg and McGrew (2019) followed up on Wineburg's initial study on historical thinking almost three decades later — after the shift from an almost exclusively offline to an online information society — to determine the skills needed to judge the credibility of online sources. Contrary to the findings on historical thinking, fact-checkers almost completely disregarded the content of online news articles to engage in a structured search on the credibility of the author and the news outlet. This strategy allowed them to identify misleading information more quickly and accurately than traditional vertical reading approaches. Unlike traditional reading strategies, which involve carefully analyzing a single document in isolation (a practice known as vertical reading), lateral reading encourages students to leave the document and seek out additional information from other sources. This method emphasizes cross-referencing information across multiple documents and websites to assess the credibility of the original source.

Wineburg et al. (2022) argue that, in times of digesting online information in an interconnected web of information, studies on multiple documents cut short on ecological validity of our everyday information environment. Unlike the multiple document paradigm, lateral readers ignore most irrelevant sources to cope with the endless stream of Internet information. Also, the goal of lateral reading is finding credible information, while the multiple source paradigm focuses on building disciplinary knowledge. However, both paradigms focus on investigating the source to analyze its credibility.

As reviewed by McGrew (2024), lateral reading interventions have proven highly effective in helping students navigate the digital landscape in formal contexts like school curricula (e.g., Brodsky, Brooks, Pavlounis, & Johnston, 2023; Wineburg et al., 2022) and informal contexts (Fendt et al., 2023).

Overall, the research fields on different source credibility approaches often overlap, have been forked off from older approaches or may

eventually merge again. Historical thinking can be seen as a foundational approach, as early work by Wineburg (1991) established core principles of sourcing, contextualization, and corroboration across multiple documents. Building on this foundation, multiple document literacy emphasizes the cognitive processes required to comprehend, integrate, and evaluate information from multiple, often conflicting, texts (Britt et al., 1999). While distinct, it shares historical thinking's core requirement of working with more than one source. Sourcing, in turn, often isolates and deepens this specific component from historical thinking, applying it to a broader range of contexts beyond history (e.g., De La Paz et al., 2014; Lescarret et al., 2024). Thus, historical thinking, multiple document literacy, and sourcing are not entirely discrete; they represent a family of related practices with a shared lineage.

In contrast, lateral reading is a more recent and distinct approach, developed specifically for navigating the online information landscape by adapting historical thinking's contextualization and sourcing principles while de-emphasizing corroboration (Wineburg et al., 2022). However, even this distinction may be blurring, as recent evidence suggests lateral reading interventions are enhanced by incorporating elements of content assessment (Fendt, Scheibenzuber, Edelsbrunner, & Nistor, 2025). For this paper, we regard lateral reading as a distinct approach that needs further research but investigate whether the other three approaches — historical thinking, sourcing, and multiple document comparison — lead to comparable or different effects.

5. Individual differences predicting source credibility assessment

Individual differences such as motivational aspects, prior knowledge, beliefs, cognitive abilities, and demographics predict variation in source credibility assessment. In a systematic review of 72 multiple document comparison studies, Anmarkrud et al. (2022) concluded that individual differences may play an important role in interventions aimed at fostering source credibility assessment. While the review highlights that participants may need basic prior knowledge to benefit from interventions, the relationships between individual differences and intervention effects vary considerably depending on the topic and the way in which skills are measured. These individual differences may influence how learners engage with source credibility assessment tasks and ultimately affect their ability to critically evaluate sources. Individuals with high personal or extrinsic motivation may be able to compensate for difficulty spikes (i.e., moments when the cognitive demands of a task temporarily and often unexpectedly increase) more easily, while learners with more prior knowledge may find it easier to familiarize themselves with new content in this area (as conceptualized by Britt, Rouet, & Durik, 2018; Panizza et al., 2022).

Demographic influences (e.g., age, gender, and educational level) on learning have also been investigated extensively (Martin, Sun, & Westine, 2020). For example, the long-standing idea that citizens who grew up in a world with digital technologies might find it easier to adapt to new technologies could not be confirmed in recent research (Peng & Yu, 2022). Furthermore, different countries might emphasize critical thinking, source evaluation, and media literacy to different extents in their national curricula. For example, in countries that focus on rote memorization, students might be less familiar with critical evaluation skills (e.g., McBride, Xiang, Wittenburg, & Shen, 2002; Nielsen, Martínez-García, & Alastor, 2021).

In their MD-TRACE model, Macedo-Rouet et al. (2013) conceptualized the role of individual differences as permanent internal sources, which include reading skills, prior knowledge on the topic, and working memory. These internal cognitive resources enable readers to process information more effectively and engage in source credibility assessment by drawing on their understanding of the subject and cognitive capabilities. Prior knowledge equips individuals with a framework to analyze information critically, allowing them to adopt a more analytic and evaluative stance, especially when it comes to scrutinizing

claims. List and Alexander (2017) further argue that prior knowledge helps readers take an analytic-critical stance by connecting new information with existing knowledge, making it easier to assess the credibility of information and sources. Limited prior knowledge might result in higher cognitive load because students would not only have to learn the respective source credibility assessment skills but also partially gather knowledge about the topic of the texts.

6. Intervention characteristics

Characteristics of the learning environment, especially the length, the learning setting, the student–instructor ratio, and digital or in-person settings, might also explain variation in intervention effectiveness.

Shorter interventions, such as one-day workshops or single-session training (e.g., Griffin et al., 2024; Lee, Moore, & Hancock, 2024; Muis, Denton, & Dubé, 2022), can be flexibly integrated into most learning settings and may be effective for introducing basic source credibility assessment approaches. This flexibility is maximized by automated interventions in which participants receive prompts, videos, or reading materials (e.g., Axelsson & Nygren, 2024; Fendt et al., 2023; Kim & Hannafin, 2016; Lee, 2021; Lescarret et al., 2024; Ulyshen, Koehler, & Gao, 2015). However, these interventions cannot provide learners with extensive topic knowledge and may need to be reinforced by follow-up practice for sustainable learning effects. Longer interventions, such as multi-week courses or year-long curricula, may offer the opportunity for more sustained and transferable learning outcomes (e.g., De La Paz, Monte-Sano, Felton, Croninger, Jackson, & Piantedosi, 2017; Nokes et al., 2007; Reisman, 2012).

Longer interventions can often be embedded more easily in structured, formal learning settings, such as in schools or universities. These contexts also often provide the opportunity for better instructional guidance (Brooks, 2011). Informal settings, for example in open online courses or youth welfare institutions, might provide increased flexibility and independence from school laws (Hoekstra, Beijgaard, Brekelmans, & Korthagen, 2007; Johnson & Majewska, 2022). Participation in informal learning activities is often voluntary, which might increase motivation, but can also result in a self-selection bias (Degner, Moser, & Lewalter, 2022). Different learning settings also often feature a diverse student–instructor ratio, affecting the amount of individual attention and feedback each student receives (Blatchford, Bassett, & Brown, 2011; Finn, Gerber, & Boyd-Zaharias, 2005). While in small-group settings, instructors can provide more personalized guidance and feedback, larger groups tend to be more cost-effective.

While most school settings are predominantly analog, digital learning environments might offer learners more flexibility to engage in self-guided learning at their own pace (Tomasik, Helbling, & Moser, 2021). Digital settings can profit from various digital technologies, such as quizzes, multimedia content (e.g., Martini et al., 2025), digital tools for information search (e.g., Stadler & Bromme, 2008), or serious games (as demonstrated by Barzilay et al., 2023). Furthermore, more open learning and testing environments provide more ecological validity, which increases the chance that learners transfer and use the skills in their everyday lives (as argued by Wineburg et al., 2022). However, digital interventions may be less effective for learners who struggle with digital literacy or who lack access to reliable Internet access and technology (Tang & Chaw, 2016).

7. The current study

The aim of the present study is to conduct a meta-analysis on interventions targeted at fostering source credibility assessment. Prior reviews have made valuable contributions to our understanding of this field, but each has specific limitations. Brante and Strømsø (2018)

provided a general discussion of source credibility assessment interventions but lacked quantitative synthesis of intervention effects. Anmarkrud et al. (2022) focused primarily on individual differences predicting source credibility assessment without systematically examining intervention effectiveness. McGrew (2024) conducted a qualitative review specifically of lateral reading interventions, but not other source credibility assessment approaches.

To the best of our knowledge, no meta-analysis of lateral reading has been conducted, nor has there been a comprehensive effort to categorize or a quantitative analysis comparing the effects of the four source credibility assessment approaches. Consequently, structured meta-analytic evidence on the overall effectiveness of source credibility interventions, as well as the moderating effects of intervention characteristics and individual differences, is still missing.

We address this gap meta-analytically and compare the effects of source credibility assessment approaches stemming from different research fields (historical thinking, sourcing, multiple document literacy, lateral reading). We aim to provide insights into which conditions work best to increase the effects of source credibility assessment interventions. As moderator variables potentially explaining variability in effect sizes, we assess the role of individual differences, (sample country, age, gender composition, educational level) and intervention characteristics (source credibility assessment approach, source-content, learning setting, intervention length, mode of delivery, digital technologies during the intervention, student–instructor ratio, intervention type, instructional support) on the intervention effectiveness, as well as several other study characteristics (type of dependent variable, type of control group, publication type, year, experimental design). We seek to answer the following questions:

RQ1: Is there an overall meta-analytic effect of source credibility assessment interventions in comparison to control conditions?

In addition, we inspect publication bias to estimate to which degree the estimated meta-analytic effect might be distorted by such bias and to estimate a bias-corrected estimate.

RQ2: How much heterogeneity do we find in the efficacy of source credibility assessment interventions?

RQ3: Which moderator variables can explain heterogeneity in effect sizes between studies?

We expect an overall moderate effect of interventions to foster source credibility assessment, in line with previous reviews (Anmarkrud et al., 2022; Brante & Strømsø, 2018; McGrew, 2024). The review by Anmarkrud et al. (2022) identifies effects of individual differences in source credibility assessment. There is also a considerable, although not specific to source credibility assessment, body of research on the effect of intervention characteristics on learning outcomes (e.g., Johnson & Majewska, 2022; Tomasik et al., 2021). We therefore expect moderating effects of individual differences and intervention characteristics on source credibility assessment interventions.

8. Method

This meta-analysis follows the PRISMA guidelines (Page et al., 2021). The pre-registration of the research questions, search strategy, data extraction, and analyses (<https://osf.io/2qatd>) as well as supplementary materials including the dataset, survey files, coding schemes, and analysis script can be found on <https://osf.io/u3bhv>. This meta-analysis did not involve direct data collection from human participants and therefore did not require approval from an ethics committee.

8.1. Search strategies

We employed three strategies to find eligible studies. An overview over the study selection procedure can be found in Fig. 1. First, we used the metasearch engine EBSCOhost to search scientific databases (Academic Search Complete, ERIC, APA PsycInfo, APA PsycArticles, OpenDissertations) for studies published before March 2025 (search

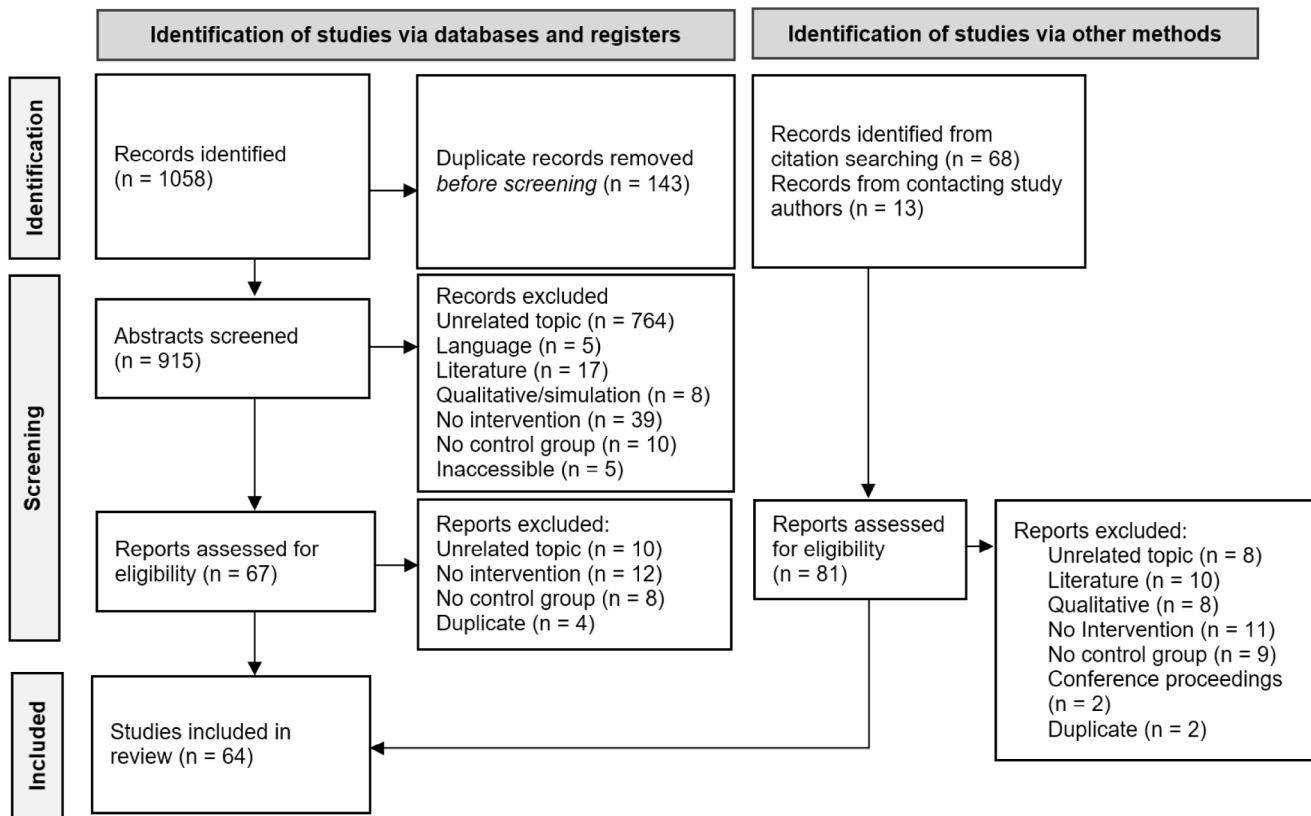


Fig. 1. Study selection procedure.

conducted on March 1, 2025). The search terms (in title, abstract, keywords or full text) were informed by reading the articles of key studies published within the different fields concerned with source credibility assessment (i.e. Anmarkrud et al., 2022; Brante & Strømsø, 2018; McGrew, 2024; Scharrer & Salmerón, 2016). The search terms were: “lateral reading” OR “read* lateral*” OR “civic online reasoning” OR sourcing OR “multiple document literacy” OR “multiple document comprehension” OR “source evaluation” OR “digital literacy intervention”. We also included “effect” OR “control group” OR experiment* to focus on controlled experimental studies. We considered the term “source credibility”, but dropped it, because most search results focused on features of credible sources, rather than fostering skills, and no additional studies met our inclusion criteria.

Second, we searched the references of literature reviews and eligible studies we found in step 1 for more eligible studies. Third, to include gray literature, we contacted each corresponding author and, if they did not provide valid contact details, the first author of each eligible study and asked for unpublished studies that met our inclusion criteria. After removing duplicates, 915 articles were included in the abstract and full-text screening.

8.2. Eligibility criteria

We selected studies based on the following criteria:

Source credibility assessment intervention. Eligible studies had to focus on fostering participants’ ability to correctly identify the credibility of sources. That is, studies that feature exclusively on assessing claims, often also called fact-checking were not included (for a meta-analysis on fact-checking interventions, we recommend Walter, Cohen, Holbert, & Morag, 2020). While other professions feature different understandings of sourcing (e.g., a way to find rich resources for production), we specifically focused on a pedagogical/psychological definition: As stated before, we define source credibility assessment as evaluating the

credibility of a text using and crosschecking available first-hand and second-hand information about the source and its meta-characteristics. The studies also had to feature an intervention that we neither restrained in length nor intensity. Thus, the interventions could be as short as nudging prompts to check source credibility (as commonly found in misinformation research), a short self-directed training, or week-long curricula.

Controlled experimental study. This meta-analysis aims at drawing causal inferences on the effect of interventions on participants’ source credibility assessment skills. The studies had to feature an experimental or quasi-experimental design with at least one treatment and one control condition. The treatment condition had to actively focus on checking the source’s credibility, whereas this was not the case in the control condition. Studies without a control group or without an intervention (e.g., studies on participants’ untrained abilities to identify information credibility) were excluded from the analysis.

Effect sizes and language. Only studies in English were considered for the analysis. Furthermore, the studies had to sufficiently report the data to compute the effect sizes (e.g., descriptive statistics with means, group sizes, and standard deviation or effect sizes).

8.3. Coding procedure

The coding scheme was developed based on our expertise and theoretical assumptions in the field and refined after every round of coder training, based on a discussion of the coders. First, the first author and three apprentice coders coded the eligibility of the studies with a 20% overlap, reaching unanimous agreement on which studies to include, but still discussing any uncertainties.

Second, for the coder training, 20% of the studies were double-coded (similar to Edelsbrunner, Simonsmeier, & Schneider, 2025; Patzl, Oberleiter, & Pietschnig, 2024), reaching an initial interrater agreement

of Gwet's AC1 = .87, 95% CI [.84, .89] (an interrater reliability index more precise and versatile than Cohen's Kappa, see Gwet, 2014). Some of the disagreements could be attributed to spelling mistakes and excluding those resulted in Gwet's AC1 = .91, 95% CI [.90, .94]. Following the best practice suggested by Hammer and Berland (2014), the coders used disagreements in their coding as indicators of where their concepts of the data misaligned to discuss rules that ensured a shared understanding.

For instance, the raters disagreed on the sample size in Axelsson and Nygren (2024) because the apprentice coders counted the overall sample, whereas the expert coder summed the included groups. This discrepancy was resolved with a rule to report the sample size as the sum of all included groups. In other cases, codes slipped into the wrong column, resulting in a rule that the expert coder would perform plausibility checks on all data and try to code any missing data. Minor issues, such as reporting a different number of digits, were solved by agreeing to always use the exact number provided by the authors of the papers. After that, the four coders coded the studies independently but discussed any uncertainties while examining the respective coding together.

8.4. Variables

We coded each dependent variable representing an outcome related to source credibility. We considered search behavior (skills to effectively search for meaningful sources), reasoning skills (the ability to logically analyze the credibility of sources/content), source knowledge (knowledge about sources and source features), and credibility judgment (judgment of the credibility of the source).

As moderators, we coded control variables (type of dependent variable, type of control group, publication type, year, experimental design), participant characteristics (sample country, age, gender composition, educational level), and intervention characteristics (source credibility assessment approach, source-content, learning setting, intervention length, mode of delivery, digital technologies during the intervention, student-instructor ratio, intervention type, instructional support).

The *type of the dependent variable* encompassed the study outcomes related to source credibility assessment described above: credibility judgment, reasoning skills, search behavior, and source knowledge. For example, Fendt, Scheibenzuber, et al. (2025) measured credibility judgment (i.e., distinguishing between credible and unreliable sources) in a source credibility test. In this test, participants rated the credibility of 3 credible and 3 untrustworthy sources. We coded this as credibility judgment. Mason, Moè, Tornatora, and Ronconi (2022) had participants write an essay about source credibility and judged whether they considered and weighed the conflicting perspectives of both provided sources and their limitations. We coded this as reasoning skills. Marten et al. (2025) checked participants' corroboration using multiple-choice questions about location, time, originator, and event in a screenshot. These questions were arranged so that participants could answer more questions, if they corroborated more thoroughly and engaged in a structured search procedure. We coded this as search behavior. Marten and Stadtler (2025) asked participants to write down all options for evaluating an imaginary claim on the Internet within 5 min and coded mentions of sourcing and corroboration. We coded this as source knowledge.

The *type of control group* was coded as instructed (e.g., regular lessons or contents other than source credibility assessment), waiting control (received intervention after undergoing pre-/post-measurements), and passive control (filling out tests but receiving no instruction).

The *publication type* was coded as article for studies published in scientific journals, thesis for work written to obtain an academic degree that was not published elsewhere, and manuscript for any study that

was not yet published. The *publication year* was coded as provided by the publisher.

The *experimental design* was coded as experimental vs. quasi-experimental, as indicated in the studies. In several studies the authors described that they assigned full classrooms to different conditions, which we coded as quasi-experimental.

The *country* was coded by the international country code. The *participant age* was extracted from the reported average participant age in the studies. The *gender composition* was calculated as the percentage of female and non-binary participants in the overall sample to include non-binary participants — a group that would have otherwise been too underrepresented in the studies for statistically meaningful analyses. For a robustness check, we also clustered non-binary participants with male participants. *Educational level* was also used as an indicator of prior knowledge and coded as primary education, secondary school qualification, university entrance qualification, university students, and graduates.

The *source credibility assessment approach* was coded as sourcing, lateral reading, multiple document literacy, and historical thinking (see section "Different approaches to source credibility assessment interventions" for definitions and detailed descriptions of the approaches). When an intervention did not explicitly mention a specific source credibility assessment approach, we analyzed the features disclosed in the paper and categorized the intervention according to our definitions.

We coded the *source-content focus* of the interventions based on their descriptions in the studies. Interventions that promoted claim investigation, corroboration, or any other form of content evaluation were classified as focusing on *both* source and content. Since a focus on *sources* was a prerequisite for inclusion in this meta-analysis, all included interventions inherently addressed source evaluation.

The *learning setting* was coded as school, university, and informal. The *length of the intervention* was coded in minutes. The *mode of delivery* was coded as digital, analog, and hybrid (i.e., a mix of the former two).

The *digital technologies* during the intervention were coded to indicate the support by digital tools during the intervention. *None* was coded if no use of digital technology was indicated. *Multimedia* was coded if the authors indicated using audio or video. *Limited websites* were either mock websites or only contained a limited set of clickable links to keep participants in a controlled environment. *E-learning platforms* (e.g., Moodle) were similar to limited websites with the addition of more interactive elements like exercises, forums, or learning journals. The *open Internet* was coded if participants could actively engage in unrestricted open web searches.

Student-instructor ratio was coded to indicate how many learners a trainer was responsible for. *High* for more than 40 students per trainer, commonly found only in massive (online) courses. *Medium* was 5 to 40 students per trainer, which would encompass most classroom settings. *low* was coded for less than 5 students per trainer, indicating a smaller ratio than classes and seminars, nearing a 1-to-1 setting. We also coded *none* for interventions that involved no human trainers, for example in a fully self-guided reading intervention.

Instructional support was coded as *static* for any type of non-individualized support, for example videos or reading materials, while *dynamic* was coded for all types of tailored support, such as human support, adaptive online environments, etc.

The *intervention type* was coded to indicate the overall idea of how the information was taught. A *curriculum* spanned multiple training sessions, usually over several weeks, and often took place in schools and universities. *One-time training* could often be found in informal settings, such as a short online course for elderly people. *Passive reception of media* would include any kind of passively watching a video, listening to an audio file, or reading a nudging prompt without cognitively activating participants by encouraging practice or providing active, tailored guidance.

8.5. Analysis

We followed the procedure described by Borenstein, Hedges, Higgins, and Rothstein (2009) for effect size calculation and moderator analysis. We gathered the data of the main statistical values and the moderators in an Excel sheet. We then used the metafor package (version 4.8, Viechtbauer, 2010) in R (version 4.5) to calculate the corrected effect sizes (Hedges g), for effect aggregation, and to implement metaregression models (Hunter & Schmidt, 2004).

To examine research question 1 (overall effectiveness), we followed two approaches. First, we calculated a three-level meta-analytic model via hierarchical linear modeling (Raudenbush & Bryk, 2002) because multiple studies reported more than one effect. In this model, we examined sampling variance at level 1 (i.e., variance due to studies only collecting samples from the entire population), level 2 (i.e., the variance τ^2 between different effects within the same study), and level 3 (i.e., the variance σ^2 between different studies). We used random-effects models in which the effect sizes could vary between the effects and studies. Following Chernikova et al. (2020), we also implemented a second model to examine the robustness of the meta-analytic effect size. Specifically, a part of the reviewed studies reported pre-test differences between intervention and control conditions in addition to post-test differences. In the second model, we only included these studies, controlling for prior differences between conditions by subtracting the pre-test differences from the post-test differences.

To address publication bias, we present a funnel plot, calculate Egger's test, and apply the trim and fill procedure. We also used PET-PEESE (Stanley & Doucouliagos, 2014) to test for small-study effects (i.e., if smaller studies report larger effect sizes). PET (Precision-Effect Test) regresses effect sizes on standard errors to test for bias, while PEESE (Precision-Effect Estimate with Standard Error) refines estimates using variance. If PET is nonsignificant, PEESE provides the corrected effect size. To identify p-hacking, we employed a z-curve analysis (Bartoš & Schimmack, 2022). This method provides information on the expected replication rate (i.e., the predicted success of exact replication studies) and expected discovery rate (i.e., the estimated proportion of the statistically significant results out of all findings). We chose this method because it provides fairly robust results for meta-analyses with considerable heterogeneity. We also used leave-one-out analyses to investigate the stability of summary effect calculations by leaving out a different effect for each recalculation and averaging the results. Substantial numerical changes in the summary effect estimates may suggest that certain effect sizes exert a distorting influence on the overall summary effect. We also calculated Cook's distance to assess the influence of individual studies on the overall results of the meta-analysis (Viechtbauer & Cheung, 2010).

To examine research question 2, which concerned heterogeneity in effect sizes, we used the Q statistic to assess whether and to what extent there was variance beyond the expected sampling errors. We also report I^2 (the estimated overall proportion of observed variation in effect sizes due to systematic study differences beyond sampling error), τ^2 (the explained variance within studies), σ^2 (the explained variance between studies), and prediction intervals to quantify how much variation would be expected in replication studies from the same overall population of studies (Edelsbrunner et al., 2025; Int'Hout, Ioannidis, Rovers, & Goeman, 2016).

Research question 3 concerns potential moderating variables of the average meta-analytic effect. We included control variables (type of dependent variable, type of control group, publication type, year, experimental design), participant characteristics (sample country, age, gender composition, educational level), and intervention characteristics (source credibility assessment approach, source-content, learning setting, intervention length, mode of delivery, digital technologies during the intervention, student-instructor ratio, intervention type, instructional support) in separate metaregression analyses. To avoid multicollinearity, we included each moderator in a separate model

with robust estimation to assess their effects separately (following the procedure of Patzl et al., 2024). We then compared the levels of categorical moderators and conducted single linear precision weight meta-regressions for continuous moderators (e.g., age, gender composition).

For statistical inference, we use a 10% Alpha error level to prevent increased Beta-/type II-error frequency, 95% confidence intervals, and 95% prediction intervals.

9. Results

We included 60 articles (see <https://osf.io/69jsa>) published between 2002 and 2025 that yielded 64 eligible studies (i.e., independent samples that neither overlapped nor shared the same control group). From these 64 studies, we extracted 246 effect estimates, representing different outcomes. The total number of participants across all studies was 17,120, of which 59.2% were female and 0.6% non-binary. The majority of the studies focused on sourcing (23), followed by multiple document literacy (20), lateral reading (14), and historical reasoning (7). Half of the studies featured adolescent participants (32), followed by adults (24), and children (8). This is reflected in the average age of the study participants, which ranged from 9.76 to 45.50 years. The samples were from the US (24), Germany (10), Israel (9), Norway (3), Italy (3), France (3), Taiwan (2), Canada (2), Finland (2), the UK (2), Argentina (1), the Netherlands (1), Spain (1), and Sweden (1) — meaning that the vast majority of studies were conducted in Western countries.

The majority featured a one-time training (31) or curricula (27), with only a few passive reception interventions (6). Most interventions took place at schools (39) or universities (20), while only few were conducted in informal learning settings (5). The average intervention length was 440.02 min ($SD = 1,065.26$, $Md = 90$, $range = 5-5400$) over an average of 4.85 intervention dates ($SD = 9.66$, $Md = 1.5$, $range = 1-60$). The control groups in most studies were instructed (41) or passive control groups (21), with only two waiting control groups. Almost all studies were published articles (59), with only few unpublished manuscripts (3), and theses (2). There were slightly more quasi-experimental (36) than experimental studies (28). Most interventions were delivered in person (44), with only a minority of digital (12) or hybrid interventions (6) and two unclear settings. Most interventions had a medium student-instructor ratio (32) or no human trainers (16), with only 6 studies featuring a high and 2 a low student-instructor ratio (student-instructor ratio in 8 studies was unclear). Most studies featured no digital technologies during the intervention (26), while some used e-learning platforms (17), the open Internet (8), limited websites (7), or multimedia (6).

9.1. Overall effect of source credibility assessment interventions

The overall estimated meta-analytic effect of the 246 effects was significant and positive, with a moderate effect size $g = 0.42$, $SE = 0.04$, $p < .001$, 95% PI $[-0.33, 1.17]$, indicating that source credibility assessment interventions lead to better outcomes compared to control conditions. The analysis indicated significant overall heterogeneity ($Q(245) = 1707.44$, $p < .001$, $I^2 = 87.45\%$) between studies ($\tau^2 = 0.02$, $SE = 0.15$) and within studies ($\sigma^2 = 0.12$, $SE = 0.35$).

Fig. 3 presents the average effect sizes of the individual studies, along with their weights, confidence intervals, and the summary effect from the random-effects model, shaded for the different source credibility approaches to illustrate these analyses.

Regarding publication bias, the funnel plot (see Fig. 2) of effect size distributions and standard errors indicated no funnel plot asymmetry. This impression was supported by a non-significant Egger's test result, $t(245) = 1.30$, $p = .193$. The trim and fill procedure imputed no effect sizes. The PET-PEESE results (Stanley & Doucouliagos, 2014) showed both a significant PET ($\beta = .62$, 95% CI $[.58, .66]$, $p < .001$) and PEESE

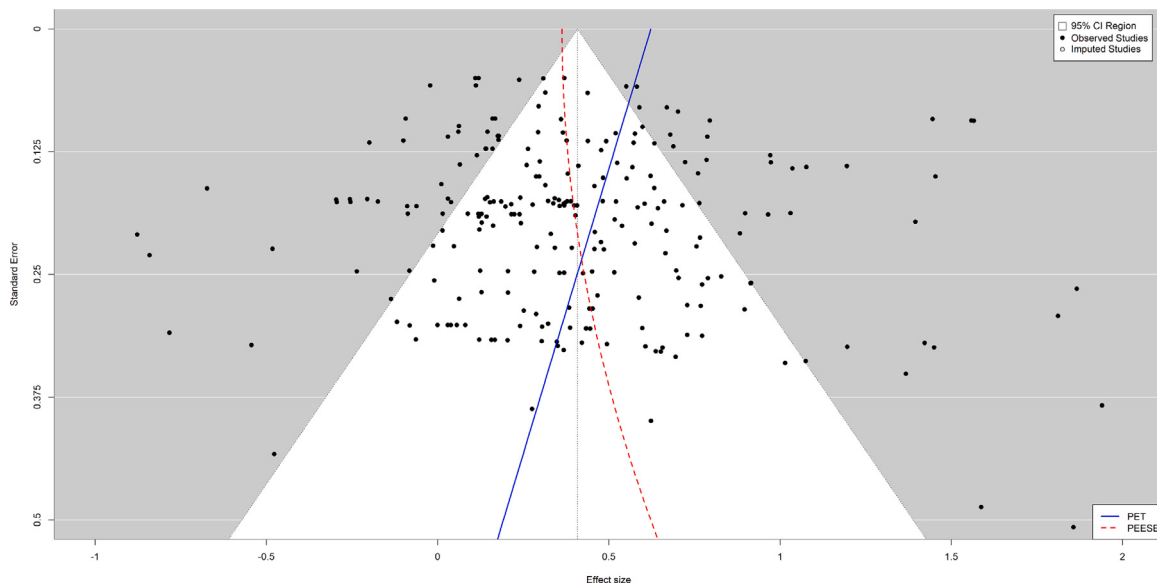


Fig. 2. Funnel plot of the effect sizes after excluding outliers.

intercept ($\beta = .36$, 95% CI [.35, .37], $p < .001$), correcting the effect size estimate to $g = 0.36$, based on the PEESE model. The z-curve analysis included 99 out of 246 valid (i.e., non-significant) effects, with an observed discovery rate (ODR) of 51%, 95% CI [.45, .57]. The expected replication rate, representing the estimated proportion of studies that would yield significant results in exact replications, was 69%, 95% CI [.55, .82]. The expected discovery rate, which estimates the proportion of true positive results among all conducted studies, was 16%, 95% CI [.06, .32]. The maximum Cook's distance of .0621 (mean = .0036) still falls below the threshold for influential studies, suggesting that there are no extreme outliers (Viechtbauer & Cheung, 2010). Overall, these results indicate that publication bias is unlikely in the field of source credibility assessment interventions.

The summary effect also remained relatively stable when individual effect sizes were omitted from analyses ($g_{mean} = 0.41$), ranging from $g = 0.40$, 95% CI [0.35, 0.46] to $g = 0.41$, 95% CI [0.36, 0.46]. Heterogeneity remained substantial across iterations ($I^2 = 86.61\% - 87.42\%$) with no single study significantly altering the overall conclusions of the analysis.

As described above, we implemented a second approach in which we corrected post-test differences between conditions for pre-test differences. This information was available for $k = 108$ effect sizes. Controlling for potential pre-test differences between the groups resulted in smaller heterogeneity ($\tau^2 = 0.08$, $\sigma^2 = 0.10$, $I^2 = 80.93\%$) and a larger overall effect, $g = 0.54$, $SE = 0.07$, $p < .001$, 95% PI [-0.30, 1.38]. A forest plot of these pre-test-corrected estimates can be found in the supplements (<https://osf.io/g3btm>).

We also calculated a model that combined the two approaches by including the pre-test-corrected value whenever possible and otherwise using the uncorrected effects. Compared to the first model, this resulted in similar heterogeneity ($\tau^2 = 0.03$, $\sigma^2 = 0.11$, $I^2 = 84.68\%$) and overall effects ($g = 0.44$, $SE = 0.04$, $p < .001$, 95% PI [-0.31, 1.19]). Trim and fill did not correct the model substantially.

9.2. Heterogeneity in effect size estimates

The meta-regression model estimated substantial heterogeneity between the studies ($Q(245) = 1707.44$, $p < .001$). The estimated between-study variance of $\tau^2 = 0.02$, $\sigma^2 = 0.12$, $I^2 = 87.45\%$ of the total variability in effect sizes could not be explained by sampling error, suggesting that the observed effects vary greatly across studies. This heterogeneity resulted in a 95% prediction interval ranging from -0.33

to 1.17, indicating that in a random study from among the population of studies reviewed, expected effect sizes range from small negative to large positive effects. Including all significant moderators (see Table 1) in a meta-regression reduced the unaccounted heterogeneity to $\tau^2 = 0.00$, $\sigma^2 = 0.12$, $I^2 = 84.88\%$.

9.3. Effects of moderator variables

A table with the coding of all moderators for each study can be found on <https://osf.io/sxzvn>. The metaregressions on publication type, publication year, and experimental design did not reveal significant influences of these variables ($p > .100$). The type of control group significantly moderated the intervention effects ($Q_M(3) = 41.73$, $p < .001$, $I^2 = 87.17\%$, $\tau^2 = 0.02$, $\sigma^2 = 0.12$), with instructed control groups yielding significantly larger effect sizes than waiting ($b = 0.11$, $SE = 0.05$, $p = .015$) and passive control groups ($b = 0.18$, $SE = 0.07$, $p = .015$).

Type of dependent variable. The type of the dependent variable also significantly moderated the intervention effects ($Q_M(4) = 41.31$, $p < .001$, $I^2 = 86.98\%$, $\tau^2 = 0.02$, $\sigma^2 = 0.12$), with search behavior ($b = 0.24$, $SE = 0.12$, $p = .044$), reasoning skills ($b = 0.15$, $SE = 0.08$, $p = .067$) and source knowledge ($b = 0.21$, $SE = 0.10$, $p = .038$) yielding significantly larger effects than credibility judgment.

9.3.1. Participant moderators

Table 1 shows the effects of the moderators in general and on their individual levels.

Country. Country significantly moderated the effects ($Q_M(14) = 13.02$, $p < .001$), reducing the heterogeneity to $I^2 = 87.08\%$, $\tau^2 = 0.01$, $\sigma^2 = 0.13$. The effects of the different countries were: Canada ($g = 0.66$, $SE = 0.68$, $p = .375$), Finland ($g = 0.28$, $SE = 0.10$, $p = .033$), France ($g = 0.22$, $SE = 0.16$, $p = .203$), Germany ($g = 0.42$, $SE = 0.06$, $p < .001$), Israel ($g = 0.38$, $SE = 0.06$, $p < .001$), Italy ($g = 0.28$, $SE = 0.08$, $p = .004$), Netherlands ($g = 0.60$, $SE = 0.14$, $p = .054$), Norway ($g = 0.36$, $SE = 0.13$, $p = .015$), Spain ($g = 0.66$, $SE = 0.32$, $p = .084$), Sweden ($g = 0.61$, $SE = 0.14$, $p = .051$), Taiwan ($g = 0.92$, $SE = 0.49$, $p = .199$), UK ($g = 0.24$, $SE = 0.06$, $p = .014$), US ($g = 0.45$, $SE = 0.06$, $p = .001$).

Age. We found no significant effect of participant age ($Q_M(1) = 0.23$, $p = .636$, $I^2 = 88.13\%$, $\tau^2 = 0.02$, $\sigma^2 = 0.12$).

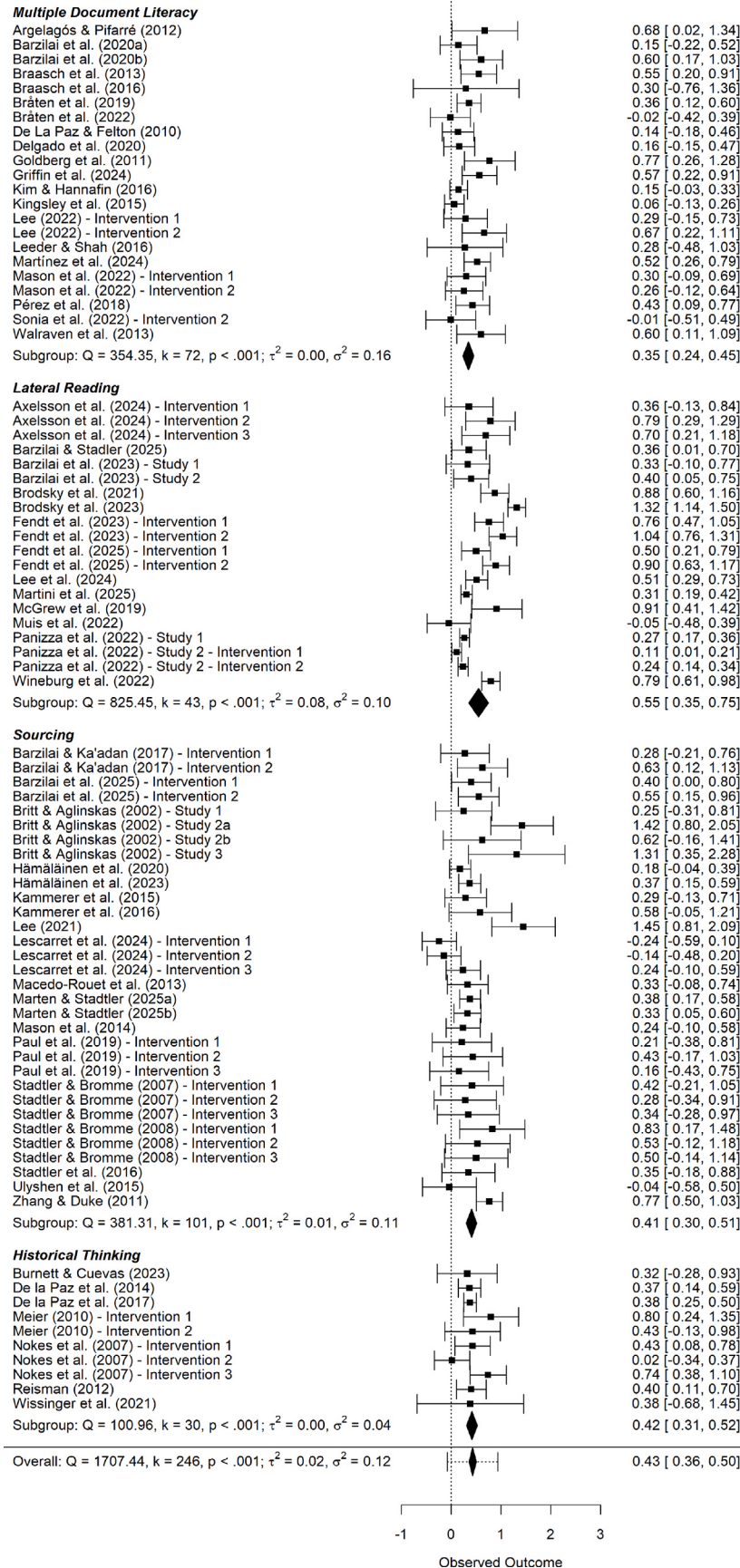


Fig. 3. Forest plots of the effect sizes for each source credibility approach.

Table 1

Effect size, confidence intervals, and prediction intervals for moderators overall and individual moderator levels.

Moderator	p of Q	Q	I^2	τ^2	σ^2	g	SE	k	95% PI	95% CI
–	<.001	1707.44	87.45	0.02	0.12	0.42	0.04	246	[−0.33, 1.17]	[0.35, 0.49]
Only studies with pre-test	<.001	551.68	80.93	0.08	0.10	0.54	0.07	108	[−0.30, 1.38]	[0.41, 0.68]
All studies (corrected for pre-test differences)	<.001	1269.83	84.68	0.03	0.11	0.44	0.04	246	[−0.31, 1.19]	[0.37, 0.52]
Country	<.001	13.02	87.08	0.01	0.13			246		
Age	.636	0.23	88.13	0.02	0.12			210		
Gender composition	.671	0.18	87.16	0.02	0.11			220		
Education	<.001	26.54	87.23	0.02	0.12			246		
Students	<.001	21.10	81.96	0.00	0.37	0.48	0.11	63	[−0.75, 1.72]	[0.27, 0.69]
University entrance qualification	<.001	27.65	61.21	0.00	0.01	0.20	0.04	21	[−0.05, 0.44]	[0.12, 0.28]
Primary education	<.001	41.39	80.94	0.01	0.07	0.41	0.06	36	[−0.18, 1.00]	[0.28, 0.54]
Secondary education	<.001	69.19	88.42	0.03	0.12	0.46	0.05	109	[−0.31, 1.23]	[0.35, 0.57]
Graduated	.001	52.68	53.81	0.00	0.02	0.78	0.11	6	[0.29, 1.26]	[0.50, 1.05]
Unspecified	.030	6.43	85.72	0.01	0.13	0.35	0.14	11	[−0.54, 1.25]	[0.04, 0.66]
Source credibility assessment approach	<.001	38.57	87.23	0.02	0.12			246		
Multiple document literacy	<.001	39.43	84.85	0.00	0.16	0.35	0.05	72	[−0.47, 1.16]	[0.24, 0.45]
Lateral reading	<.001	31.67	95.36	0.08	0.10	0.55	0.10	43	[−0.34, 1.44]	[0.35, 0.75]
Sourcing	<.001	58.47	78.13	0.01	0.11	0.41	0.05	101	[−0.30, 1.11]	[0.30, 0.51]
Historical reasoning	<.001	62.92	65.66	0.00	0.04	0.42	0.05	30	[0.01, 0.82]	[0.31, 0.52]
Student–instructor ratio	<.001	33.18	87.23	0.02	0.12			246		
Low	.014	69.18	49.48	0.00	0.02	0.94	0.11	3	[0.17, 1.71]	[0.45, 1.43]
Medium	<.001	84.10	84.79	0.04	0.10	0.45	0.05	138	[−0.27, 1.18]	[0.36, 0.55]
High	.040	5.11	93.02	0.01	0.27	0.35	0.15	15	[−0.83, 1.53]	[0.02, 0.68]
None	<.001	34.72	92.52	0.00	0.18	0.46	0.08	55	[−0.41, 1.32]	[0.30, 0.61]
Unclear	<.001	20.06	74.22	0.00	0.08	0.26	0.06	35	[−0.33, 0.86]	[0.14, 0.38]
Intervention delivery	<.001	50.39	86.27	0.01	0.12			246		
Hybrid	<.001	20.35	94.22	0.06	0.26	0.78	0.17	20	[−0.45, 2.02]	[0.42, 1.14]
In-person	<.001	137.09	81.69	0.00	0.10	0.39	0.03	175	[−0.26, 1.04]	[0.33, 0.46]
Digital	<.001	31.39	89.13	0.00	0.11	0.36	0.06	42	[−0.32, 1.04]	[0.23, 0.49]
Unclear	.360	0.94	88.56	0.00	0.33	0.24	0.24	9	[−1.20, 1.67]	[−0.33, 0.80]
Intervention type	<.001	49.03	87.47	0.02	0.12			246		
Curriculum	<.001	73.94	87.54	0.04	0.10	0.46	0.05	107	[−0.28, 1.20]	[0.36, 0.57]
One-time	<.001	61.43	84.30	0.00	0.15	0.38	0.05	122	[−0.41, 1.17]	[0.28, 0.47]
Passive	.009	8.80	93.85	0.02	0.12	0.37	0.13	17	[−0.44, 1.19]	[0.11, 0.64]
Learning setting	<.001	46.48	87.60	0.03	0.12			246		
School	<.001	108.85	84.98	0.02	0.10	0.42	0.04	167	[−0.28, 1.12]	[0.34, 0.50]
University	<.001	26.37	86.88	0.05	0.21	0.49	0.10	65	[−0.54, 1.53]	[0.30, 0.69]
Informal	.010	9.09	94.48	0.01	0.09	0.31	0.10	14	[−0.41, 1.02]	[0.09, 0.53]
Digital technologies	<.001	29.94	87.48	0.02	0.12			246		
E-learning platform (e.g., Moodle)	<.001	39.37	88.28	0.00	0.23	0.44	0.07	66	[−0.52, 1.40]	[0.30, 0.58]
None	<.001	88.36	75.78	0.00	0.10	0.38	0.04	102	[−0.25, 1.00]	[0.30, 0.45]
Open Internet	<.001	30.22	88.47	0.02	0.12	0.55	0.10	22	[−0.25, 1.35]	[0.34, 0.76]
Limited websites	.009	7.80	95.64	0.15	0.05	0.45	0.16	31	[−0.54, 1.44]	[0.12, 0.78]
Multimedia	.028	5.49	90.05	0.07	0.07	0.33	0.14	25	[−0.50, 1.16]	[0.04, 0.62]
Dependent variable	<.001	41.31	86.98	0.02	0.12			246		
Search behavior	<.001	34.41	93.53	0.07	0.14	0.51	0.09	49	[−0.42, 1.43]	[0.33, 0.68]
Reasoning skills	<.001	62.83	85.74	0.00	0.15	0.42	0.05	82	[−0.36, 1.19]	[0.31, 0.52]
Source knowledge	<.001	29.95	79.49	0.10	0.04	0.52	0.10	38	[−0.27, 1.31]	[0.33, 0.72]
Credibility judgment	<.001	29.74	80.31	0.07	0.01	0.29	0.05	77	[−0.28, 0.87]	[0.19, 0.40]
Control type	<.001	51.94	87.17	0.02	0.12			246		
Classic control	<.001	41.73	77.96	0.01	0.08	0.31	0.05	94	[−0.28, 0.90]	[0.22, 0.41]
Instructed control	<.001	102.03	89.85	0.02	0.15	0.48	0.05	141	[−0.34, 1.31]	[0.39, 0.58]
Waiting control	<.001	27.35	65.76	0.00	0.03	0.37	0.07	11	[−0.06, 0.80]	[0.21, 0.53]

Note. Results from three-level meta-analytic models consisting only of the intercept representing the overall effect of each subgroup.

Models were estimated using REML. Country moderator levels are in the respective paragraph. g = Prediction estimate. SE = Standard error.

CI = Confidence interval. PI = Prediction interval. Heterogeneity indicators I^2 (total), τ^2 (within-study), σ^2 (between-studies).

Gender composition. We found no significant effect of gender composition, neither for the contrast between male and female/non-binary participants ($Q_M(1) = 0.18, p = .971, I^2 = 87.16\%, \tau^2 = 0.02, \sigma^2 = 0.12$), nor for the contrast between female and male/non-binary participants ($Q_M(1) = 0.63, p = .429, I^2 = 87.11\%, \tau^2 = 0.02, \sigma^2 = 0.11$).

Educational level. Educational level significantly moderated the intervention effects ($Q_M(6) = 26.54, p < .001$), reducing the heterogeneity to $I^2 = 87.23\%, \tau^2 = 0.02, \sigma^2 = 0.12$. Graduated participants showed larger effects than primary education ($b = 0.35, SE = 0.11, p = .001$), secondary education, ($b = 0.34, SE = 0.11, p = .002$), university students ($b = 0.32, SE = 0.12, p = .007$), and university entrance qualification ($b = 0.59, SE = 0.10, p < .001$). Secondary education ($b = 0.24, SE = 0.07, p < .001$) and university students ($b = 0.26, SE = 0.08, p = .001$) showed larger effects than university entrance qualification. Primary education showed larger effects than participants with a university

entrance qualification ($b = 0.23, SE = 0.07, p = .001$). The other comparisons did not differ significantly, $p > .100$.

9.3.2. Intervention moderators

Table 1 shows the effects of the moderators in general and at their individual levels.

Source credibility assessment approach. The source credibility assessment approach significantly moderated the intervention effects ($Q_M(4) = 38.57, p < .001$) as indicated by a reduced heterogeneity of $I^2 = 87.23\%, \tau^2 = 0.02, \sigma^2 = 0.12$ (see Fig. 3). Lateral reading ($b = 0.21, SE = 0.11, p = .067$) yielded significantly larger effects than multiple document literacy, while the other comparisons resulted in non-significant effects, $p > .100$. A contrast showed that lateral reading studies yielded larger effects than the combined other source credibility assessment approaches ($b = 0.18, SE = 0.11, p = .093$).

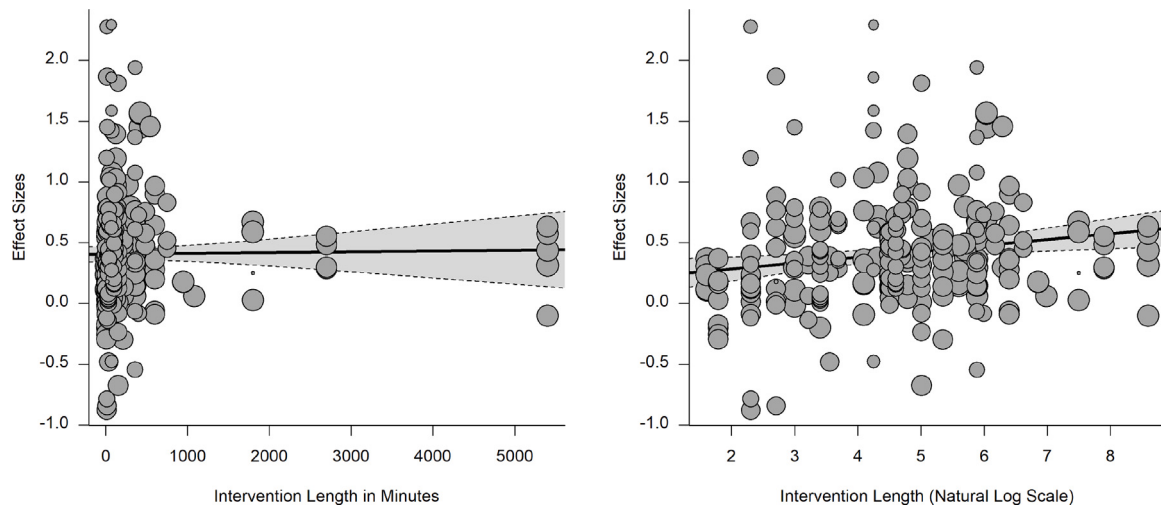


Fig. 4. Bubble plot of intervention length (in minutes and logarithmized).

A robustness check revealed a significant interaction between the source credibility assessment approach and intervention length ($Q_M(7) = 4.67, p < .001$). Compared to multiple document comparison, intervention length showed a positive interaction with lateral reading ($g = 0.001, SE = 0.00, p = .002$), and a negative interaction with sourcing ($g = -0.001, SE = 0.00, p = .068$) and historical reasoning ($g = -0.000, SE = 0.00, p = .079$). These results imply that for an additional 100 min of intervention length, the effect size of lateral reading interventions increases by about $g = 0.10$, whereas it decreases by about $g = 0.10$ for sourcing and (less so) for historical reasoning interventions.

Student–instructor ratio. Student–instructor ratio significantly moderated the effect of source credibility assessment interventions ($Q_M(4) = 33.18, p < .001, I^2 = 87.23\%, \tau^2 = 0.02, \sigma^2 = 0.12$). A low student–instructor ratio yielded larger effects than a high ($b = 0.58, SE = 0.15, p < .001$) and medium ratio ($b = 0.52, SE = 0.08, p < .001$) as well as no human trainers ($b = 0.49, SE = 0.08, p < .001$). The effect for an unclear student–instructor ratio was smaller than for a low ($b = 0.68, SE = 0.09, p < .001$) and medium ratio ($b = 0.15, SE = 0.08, p = .053$) as well as no human trainers ($b = 0.18, SE = 0.08, p = .024$). The pairwise comparisons revealed no other significant differences, $p > .100$.

Instructional support. Instructional support significantly moderated the effect of source credibility assessment interventions, ($Q_M(2) = 49.49, p < .001, I^2 = 87.72\%, \tau^2 = 0.02, \sigma^2 = 0.13$). The pairwise comparisons revealed no significant difference, $p > .100$, and dynamic support ($g = 0.45, SE = 0.05$) yielded only slightly larger descriptive effects than static support ($g = 0.41, SE = 0.05$).

Intervention delivery. The intervention delivery moderated the effect of the interventions ($Q_M(4) = 50.39, p < .001, I^2 = 86.27\%, \tau^2 = 0.01, \sigma^2 = 0.12$). Hybrid interventions showed significantly larger effects than in-person ($b = 0.43, SE = 0.17, p = .011$), digital ($b = 0.42, SE = 0.17, p = .015$), and unclear interventions ($b = 0.49, SE = 0.28, p = .084$), while the other comparisons did not differ significantly, $p > .100$.

Intervention type. The intervention type also moderated the effects, $Q_M(3) = 49.03, p < .001, I^2 = 87.47\%, \tau^2 = 0.02, \sigma^2 = 0.12$. While pairwise comparisons showed no significant results, $p > .100$, descriptively curricula ($g = 0.46, SE = 0.05$) yielded the largest effects.

Learning setting. The learning setting significantly moderated intervention effects ($Q_M(3) = 46.48, p < .001$), reducing unexplained heterogeneity to $I^2 = 87.60\%, \tau^2 = 0.03, \sigma^2 = 0.12$. Pairwise comparisons indicated that informal settings had a significantly smaller effect than

school settings ($b = -0.49, SE = 0.17, p = .004$) and university settings ($b = -0.67, SE = 0.18, p < .001$), while the difference between university and school settings did not reach significance ($b = 0.18, SE = 0.11, p = .105$).

Intervention length. Intervention length did not significantly moderate the intervention effects ($b = -.00, SE = 0.00, p = .977$). Descriptively, the longest intervention type, curricula ($M = 993.89$ min, $SD = 1,486.42$ min) showed larger effects than one-time interventions ($M = 63.82$ min, $SD = 48.36$ min) and passive multimedia reception ($M = 16.67$ min, $SD = 10.27$ min). A bubble plot (see Fig. 4) indicated a left-skewing of effect sizes, with most interventions shorter than two hours. In a second analysis, we therefore log-transformed the intervention length to account for the unequal distribution (as suggested by Ranger, Kuhn, & Ortner, 2020), which resulted in a significant effect ($Q_M(1) = 3.77, b = .04, SE = 0.02, p = .053, I^2 = 87.14\%, \tau^2 = 0.02, \sigma^2 = 0.12$).

Interaction between training type and intervention length. A model that included intervention length, training type, and their interaction did not significantly moderate intervention effects ($Q_M(6) = 1.33, p = .254, I^2 = 87.27\%, \tau^2 = 0.02, \sigma^2 = 0.12$).

Digital technologies during the intervention. The inclusion of digital technologies during the intervention significantly moderated intervention effects ($Q_M(5) = 29.94, p < .001, I^2 = 87.48\%, \tau^2 = 0.02, \sigma^2 = 0.12$). Pairwise comparisons revealed that open Internet resources had a significantly stronger effect than using no digital technology ($b = 0.19, SE = 0.10, p = .066$), while no significant differences were found between other pairs, $p > .100$.

Source-content focus. Source-content focus significantly moderated intervention effects ($Q_M(2) = 78.64, p < .001, I^2 = 85.07\%, \tau^2 = 0.00, \sigma^2 = 0.11$). While pairwise comparisons showed no significant results, $p > .100$, an intervention focus only on sources ($g = .48, SE = 0.10$) showed descriptively larger results than focusing on both content and source ($g = .37, SE = 0.12$).

10. Discussion

In our information-driven society, today's online environments — overloaded with information of often elusive credibility — pose substantial challenges, with a critical need to foster individuals' skills to detect reliable sources and trustworthy information. While research on interventions to foster source credibility assessment in an analog world is well-established and spans more than three decades (Brante & Strømso, 2018), societal developments such as the open Internet require new approaches to evaluating information (Wineburg, 2024).

Our preregistered meta-analysis showed that interventions to foster source credibility assessment have small to moderate overall effects across a range of different contexts and learning settings. To our knowledge, we provide the first analysis to include all four approaches (i.e., historical thinking, sourcing, multiple document literacy, and lateral reading), showing that source credibility assessment addressed through lateral reading interventions yielded considerably larger, moderate effects than the other strategies.

The effectiveness of source credibility assessment interventions, however, highly depends on many complex factors, especially individual differences and intervention design. At the individual level, the interventions were effective regardless of participants' age and gender composition. However, educational level, which may be regarded as an indicator for prior knowledge, played an important role: Interventions targeting high school graduates yielded smaller effects, while those involving university graduates showed the largest gains. At the intervention level, a low and medium student–instructor ratio, longer curricula, university settings, and interventions that used the open Internet provided larger effect sizes. We found that the interventions targeted different facets of source evaluation (i.e., search behavior, reasoning skills, source knowledge, credibility judgment) with comparable effectiveness.

10.1. Overall meta-analytic effect

We conclude that source credibility assessment interventions are small-to-moderately effective overall. These effects are slightly smaller than those reported in structural observations by Brante and Strömsö (2018), who found moderate overall effects. Several factors may explain these discrepancies. The high heterogeneity observed in our analysis suggests the influence of additional variables, as commonly observed in longer interventions and those with greater ecological validity. This means source credibility assessment interventions may be particularly sensitive to individual differences and intervention characteristics.

Furthermore, the overall effect increased notably when we included only studies that allowed controlling for pre-test differences. This increase might be explained by the larger scale and methodological rigor of these studies, which can be a proxy for interventions with more resources. It may also hint at pre-test differences between the groups in general (e.g., in De La Paz & Felton, 2010; Stadler, Scharrer, Macedo-Rouet, Rouet, & Bromme, 2016), which could indicate a selection bias or non-perfect randomization. Threats to randomization occur, for example, when participant data are removed (e.g., outliers or partially missing data), which can invalidate randomization because it may lead to excluding participants with specific characteristics (Rubin, 1976). While randomization at the participant level is often not feasible in more ecologically valid classroom interventions where whole classes are assigned to different conditions, we found no significant difference between experimental and quasi-experimental studies. Nonetheless, in future research, authors should indicate how they randomized learners to different conditions, whether they excluded any data, and whether other factors may have compromised randomization and the possibility of causal inference (Imbens, 2024).

10.2. Publication bias

Most of our publication bias analyses (funnel plot, z curve, trim-and-fill) did not indicate any signs of publication bias, such as fewer studies with small effects than expected. The PET-PEESE method, in contrast, yielded a small downward correction of the meta-analytic effect size, indicating some bias. While these results may appear plausible given that the overall meta-analytic effects were only small to medium, methods that correct for publication bias have severe limitations and should be interpreted with caution (Carter, Schönbrodt, Gervais, & Hilgard, 2019). We therefore recommend using pre-registration and the publication of registered reports to further minimize the risk of this type of bias in future studies.

10.3. Heterogeneity in effect estimates

In general, we found high heterogeneity estimates in the meta-analytic models. The prediction intervals were broad. For the overall meta-analytic effect, if we were to design a new study from the same population of studies as reviewed here, we would expect its effect size to fall within the range of a small negative to a large positive effect most of the time. In contrast to the positive meta-analytic effect estimate, this prediction may seem unsatisfactory. Yet, first and foremost, it indicates that there is vast space to examine moderators of intervention effects to inform theories and tailor interventions that harness this systematic heterogeneity to achieve higher effects (Tipton et al., 2023). We have taken a first step in this regard by examining various moderators on the level of the participants, interventions, and other study characteristics.

10.4. Moderator variables

We found that the three established approaches — historical thinking, sourcing, and multiple document literacy — yielded similar, small effects. As established in the introduction, these approaches are often interconnected and share many similarities, with the initial conceptions (Britt et al., 1999) and interventions (Britt & Aglinskias, 2002) on multiple document literacy heavily drawing on historical thinking, while Wineburg (1991) already included sourcing. While sourcing can be regarded as a distinct approach that was expanded after the initial conceptions in historical thinking, source evaluation is often taught with multiple documents (e.g., Lee, 2021). Synthesizing information across multiple sources can be cognitively demanding, but fosters deeper engagement with content and context, potentially enhancing evaluative judgments in the long run (Rouet & Britt, 2011). This resulted in numerous effective interventions from each strand (e.g., Axelsson & Nygren, 2024; Goldberg, Schwarz, & Porat, 2011; Hämäläinen et al., 2023; Reisman, 2012). While interventions based on these approaches are effective, they may require deeper changes to transfer to our current information environment — as demonstrated by Delgado et al. (2020), Lee et al. (2024) and Alon, Rahimi, and Tahar (2024). Similar to the conceptions of Wineburg et al. (2022), future studies could profit from meta-heuristics so that learners know when to apply a specific strategy. These heuristics could outline when people can disregard the claims because of the amount of information and solely focus on the source or when judging contextual clues is more promising than judging the source. Furthermore, interventions may need to also teach necessary basic skills on navigating the Internet and understanding its structure (as conceptualized by Wineburg & McGrew, 2019).

Lateral reading interventions, by contrast, yielded overall larger, medium effects than interventions based on the other three approaches, which we attribute mainly to three reasons. First, lateral reading might be better suited for the nature of online information contexts by encouraging verification through external, reliable sources rather than relying on surface- and content-level evaluation of information. Nonetheless, lateral reading interventions may still benefit from teaching content knowledge, especially for less familiar or ambiguous sources (Barzilai, Thomm, & Shlomi-Elou, 2020; Fendt, Scheibenzuber, et al., 2025).

Second, lateral reading counter-intuitively requires less reading than the other source credibility assessment approaches (Kozyreva et al., 2023). This may benefit people with limited frustration tolerance or lower literacy skills because a verdict can be reached quite fast. Although all approaches share foundational principles — such as evaluating the author's intent and purpose, originally suggested in historical reasoning and subsequently adapted in multiple document literacy — they may require prolonged practice and guidance to be effective in modern digital information environments. In contrast, laterally checking a source can arguably be done in a minute by experienced users and taught successfully in interventions as short as 15 min (Fendt et al., 2023). However, it should be noted that all source credibility assessment approaches, especially lateral reading, require extensive

knowledge. In particular, lateral reading emphasizes search skills that may be unfamiliar and difficult to acquire for many citizens (McGrew & Byrne, 2020).

Third, lateral reading specifically focuses on digital information and aligns more closely with individuals' real information environment and actual information-seeking practices (McGrew, Smith, Breakstone, Ortega, & Wineburg, 2019). This relevance may increase intrinsic motivation, as individuals are more likely to see the immediate practical benefit of the intervention in their everyday lives (Fendt, Hufendiek, Oberparleiter, Scheibenzuber, & Edelsbrunner, 2025). This motivational effect might be further increased by gamified interventions (Barzilai et al., 2023; Barzilai & Stadler, 2025).

Overall, while historical thinking, sourcing and multiple document literacy can be seen as distinct approaches, there is considerable overlap, with lateral reading arguably being the most distinct. However, lateral reading may still benefit from incorporating insights from the other approaches, for example content assessment (Fendt, Scheibenzuber, et al., 2025) — which may challenge its distinction.

While the source credibility assessment approach in general plays an important role, individual differences also shape how learners engage with and benefit from these strategies (Anmarkrud et al., 2014). These individual differences require different focuses and adaptations of interventions, which often poses substantial challenges for their design. We found that the interventions were similarly effective regardless of the gender composition, which is consistent with recent research on analog and digital learning (Geist & King, 2008; González-Gómez, Guardiola, Martín Rodríguez, & Montero Alonso, 2012; Hou, Nguyen, Richey, & McLaren, 2020).

We also found that source credibility assessment interventions were similarly effective among all age groups. This finding is not unexpected for offline settings and, contrary to popular belief, also extends to older people in digital settings (e.g., Lee, 2022; Stadler & Bromme, 2007). This finding can also indicate that all age groups require similar support to learn the digital literacy skills required for lateral reading. Moore and Hancock (2022) successfully integrated such basics into a lateral reading curriculum, showing that motivated learners can acquire these skills regardless of age. Nevertheless, we are not aware of any controlled experimental lateral reading studies for younger children — an age group that may lack the necessary digital literacy skills. Despite the vast body of research on literacy interventions in digital settings (e.g., Niklas, Birtwistle, Wirth, Schiele, & Mues, 2022) and source credibility assessment interventions for elementary (e.g., Barzilai, Tal-Savir, Abed, Mor-Hagani, & Chinn, 2025; Burnett & Cuevas, 2023; Macedo-Rouet et al., 2013; Paul, Stadler, & Bromme, 2019; Wissinger, De La Paz, & Jackson, 2021) and secondary school children (e.g., Argelagós & Pifarré, 2012; Braasch, Bråten, Strømsø, Anmarkrud, & Ferguson, 2013; Kammerer, Meier, & Stahl, 2016; Kingsley, Cassady, & Tancock, 2015), lateral reading interventions have yet to be successfully adapted to that age group.

Educational level played a significant role in intervention effectiveness, whereas age did not. While education may serve as an indicator of prior knowledge (as interpreted by Chernikova et al., 2020), other explanations are also possible. Higher educational attainment may be associated with stronger analytical reasoning skills, greater exposure to academic literacy, or more experience navigating complex informational environments. Such characteristics may, in turn, improve participants' ability to benefit from interventions, particularly those targeting rather complex skills. It may also imply that the participants have dropped out of formal education or have been working for several years (e.g., in Kammerer et al., 2015). In both cases, the participants may not be used to learning daily anymore or may have already struggled with learning in the first place. Conversely, age did not significantly moderate intervention effects. However, the spread of ages in our dataset was relatively narrow, primarily including participants between 10 and 30, which may have limited our ability to detect potential age-related differences.

Bridging the gap between participant and intervention characteristics, this interpretation of the effect of the educational level is further supported by the effect of learning settings, with universities and schools having the largest effects (e.g., Fendt et al., 2023; Leeder & Shah, 2016). This may be an indication of the effective learning culture in educational institutions, which often require self-directed learning, extensive reading, and allow for repeated practice throughout the academic/school year.

Longer curricula proved effective, potentially suggesting that repeated practice and the guided structure of such interventions can support sustained learning success. Naturally, these interventions also allow for a more detailed exploration of the basics, such as explaining the structure of source types in a study by Brodsky et al. (2021) or deeply engaging in historical reading strategies in a study by De La Paz et al. (2014). These settings also allow for a focus on complex content and flexibly teaching knowledge about different situations as a blueprint that can be used in unfamiliar situations. It may also point to the importance of embedding source credibility assessment skills in extended instructional contexts so that students can acquire these crucial skills at an early age (as demonstrated by Wissinger et al., 2021). Counterintuitively, passive reception of multimedia yielded similarly large effects. This may be due to the more controlled nature of such lab-like experiments, which focus on delivering a small amount of information in a short time frame, often assessing effects immediately after the intervention (Capewell et al., 2024). Furthermore, longer curricula introduce more interfering variables and a greater time lapse between the intervention and the post-test, which may influence outcomes (e.g., Brodsky et al., 2023).

Curricular settings naturally hint at a medium student–instructor ratio, which yielded results comparable to interventions with no instructors, falling short only of those with a low student–instructor ratio (e.g., Mason, Junyent, & Tornatora, 2014; Meier, 2010; Sonia et al., 2022; Walraven, Brand-Gruwel, & Boshuizen, 2013). A balanced, moderate student–instructor ratio may provide the optimal mix of individual attention and peer interaction, potentially supported by collaborative learning, while allowing instructors to support a manageable number of students. Contrary to recent findings (Tipton et al., 2023), active control groups had larger effects than passive control groups. This may be explained by confounding variables — for example, the studies with passive control groups may have been conducted under otherwise uncontrolled conditions (e.g., in classrooms). Still, this finding should be investigated further in future studies and meta-analyses.

Looking at the degree of digitization of the intervention, hybrid interventions (e.g., Fendt, Scheibenzuber, et al., 2025) proved more effective. This suggests that well-designed digital interventions employing a hybrid approach can effectively substitute for traditional face-to-face instruction. We conjecture that digital or hybrid interventions may work, particularly for learners with strong self-regulation skills. Digital interventions may allow for more flexible, self-paced learning experiences, which may be advantageous for advanced learners who can navigate digital tools independently. Although technological tools allow for responding to learner dynamics (e.g., Aleven & Koedinger, 2002), face-to-face interventions may often allow for more immediate, direct, and appropriate guidance, which may benefit learners who need additional support.

Using digital technologies to support the interventions can enhance intervention effects, especially the open Internet, but descriptively also limited websites and e-learning platforms. We regard this as an indicator of higher ecological validity (as argued by Wineburg et al., 2022), which on the one hand may make the interventions harder to control. On the other hand, the open Internet may more closely mimic the information environment that participants experience at home, increasing the chance that learners transfer intervention content to their everyday lives (e.g., Leeder & Shah, 2016; Zhang & Duke, 2011).

Dynamic and static instructional support yielded similarly large effects. This is unexpected, given that recent research found different

effects in other educational settings (e.g., [Dyrvold & Bergvall, 2023](#)). Also, [Fendt et al. \(2023\)](#), [Fendt, Scheibenzuber, et al. \(2025\)](#) directly compared a lateral reading intervention with static support to one with dynamic support. They found larger effects for dynamic support, which they attribute to the more tailored guidance provided by the trainers and to potential motivation boosts. However, most included source credibility assessment interventions naturally provide some form of instructional support. Most of the studies also lacked detailed information about the nature and quality of the instructional support. For example, most school interventions may have provided dynamic support through the presence of a teacher, but this was not reported in the paper.

Another way to provide automated dynamic support in educational settings is through large language models (LLMs), a type of machine learning/artificial intelligence (AI) and prevalent societal phenomenon ([Holmes & Tuomi, 2022](#)). However, most educational research on LLMs is still fairly new and while some studies show promising results, most are still limited by a lack of objective measures of learning improvement, no gain, or sometimes even detrimental results ([Elstad, 2024](#); [Muehlhoff & Henningsen, 2024](#)). None of the studies we included used LLMs, despite first advances in that field ([Zhang & Pradeep, 2023](#)). One undoubtedly outstanding example in a related field is a larger intervention by [Costello, Pennycook, and Rand \(2024\)](#) that employed ChatGPT-4o to successfully reduce belief in conspiracy theories. There, the LLM served as a discussion partner to persuade conspiracy believers based on well-checked claims with decades of knowledge, but did not check sources. Similarly, future interventions could employ LLMs as a dynamic support tool on well-checked claims and sources, to provide tailored feedback based on a prepared knowledge base, or to summarize complicated scientific circumstances in plain language.

10.5. Limitations

This meta-analysis carefully considered and analyzed a larger body of research, with a particular focus on individual differences and intervention characteristics. This necessarily limits the depth and number of variables we could include, based on the data the studies provided.

Notably, we did not examine motivation, which might impact the effectiveness of interventions. This factor, while only assessed in few of the included studies, may still have the potential to gain a more profound understanding of the intervention effectiveness. Furthermore, most studies did not assess participants' knowledge in a pre-test, making it difficult to confirm equal baseline knowledge between the study groups. Also, few studies included delayed post-tests (except for [Bråten, Brante, & Strømso, 2019](#); [Brodsky et al., 2023](#); [Martínez, Saux, Londra, & Burin, 2024](#)), limiting our ability to assess the long-term sustainability of source credibility assessment interventions. Furthermore, many of the reviewed studies provided little detail on the instructional approaches that would have allowed coding further characteristics, such as whether the studies employed inquiry learning settings. While we have no reason to suspect intentional under-reporting in our included studies, we cannot rule out the possibility that moderators like the source credibility approach or the use of digital technologies are confounded with unreported design features of the interventions. In future research, we recommend reporting as many details about the instructional design as possible to support replication studies and meta-analytic coding.

Similarly, we were not able to include prior beliefs — one of the main influences on judging source and information credibility ([Sultan et al., 2024](#)) — because only two studies collected political opinion ([Fendt et al., 2023](#); [Fendt, Scheibenzuber, et al., 2025](#)), two epistemic beliefs ([Lee, 2022](#); [Ulyshen et al., 2015](#)), and one trust in scientists ([Martini et al., 2025](#)). While we could not include this moderator, future studies should consider including especially prior beliefs on politics, science, or the intervention content.

Several inclusion decisions, while carefully considered, may further limit the generalizability of this meta-analysis. We included only controlled experimental studies to focus on high-quality research, excluding non-controlled studies that might have provided additional insights, for example more detailed description and assessment of the intervention group (e.g., [Abed & Barzilai, 2023](#)). Our focus on English-language studies also means our findings primarily reflect interventions in Western contexts, which may not fully represent source credibility assessment interventions in non-Western contexts. Additionally, most studies took place in formal educational environments, excluding both older age groups and individuals outside traditional education systems.

We also noticed several limitations caused by the quality of the data reported by the publications. The range of outcomes highlights a need for more standardized assessment of source credibility assessment skills. Outcomes should clearly distinguish between context, content, and source (similar to [Forzani, 2020](#)). We particularly noticed that some studies assessed vastly different outcomes — for example faithfully executing the source credibility assessment strategy and successfully drawing the correct conclusion on information credibility — in the same variable (e.g., [Brodsky et al., 2023](#)). This meant that in some of these designs, participants would receive the lowest possible score for not using the source credibility assessment strategy, even if they had reached the correct conclusion. Furthermore, several studies did not clearly report, whether the design was experimental or quasi-experimental, making it more difficult to test for this moderator.

We also noticed that the different source credibility approaches are closely related and difficult to discern in several studies, with historically diverging, parallel, or eventually merging research strands. Especially, studies that incorporated elements of multiple approaches may leave classification to the researcher's interpretation. This might have deteriorated selectivity, especially for the classification of sourcing and multiple document literacy. However, the intervention approaches of historical thinking as the primary, very context-specific source credibility assessment intervention ([Wineburg, 1991](#)), and lateral reading ([Wineburg & McGrew, 2019](#)) as the most recent approach, specifically for online information environments are more selective. Consequently, this limitation is not harmful to the inference of this meta-analysis that lateral reading yields better results than more traditional source credibility assessment approaches.

Finally, we want to highlight an additional limitation: Although the meta-analysis was conducted successfully, we were unable to carry out a planned re-analysis of the original datasets. To obtain the necessary datasets and statistically validate the results, we contacted the corresponding authors of each included study, and in cases where emails were undeliverable, we reached out to other co-authors. However, we received responses from less than 20% of the authors. While all replying researchers were supportive of our study, we were only able to obtain 4 additional datasets (10 datasets from 6 research groups were already publicly available). One additional group had promised to provide the data, but was unable to do so within 6 months. While some authors were not allowed to share their data due to ethical constraints, others reported that they had lost their data. We have no reason to doubt the sincerity of these responses, but are nevertheless concerned about the current state of open science practices in this field of research, which are fundamental to fostering sustainable, transparent, and reproducible research.

11. Implications for practice and research

The findings of this meta-analysis offer several implications for educational research and practice. For research, the greater effectiveness of lateral reading over more traditional source credibility assessment concepts suggests a need for expanded research on this approach to analyze its effects in detail. In particular, there is limited research on the role of motivational factors in lateral reading interventions. In addition, although [Wineburg et al. \(2022\)](#) conceptualized that lateral

reading is almost agnostic to content knowledge, this assumption still needs to be empirically tested. Furthermore, while we were able to analyze several individual differences that were highlighted by Anmarkrud et al. (2022), only few studies assessed the effects of motivation, interest, or epistemic beliefs at a level that would have allowed their inclusion as moderators. Future studies could assess these variables with validated scales, which are already common in educational research. The assessment of intervention effectiveness also varied considerably. Some interventions seemed to aim at vastly different outcomes but combined them in a single variable. Future interventions could aim for establishing standardized, evaluated procedures or frameworks for assessing search strategies and their success separately. Furthermore, while especially educational settings and longer curricula would be ideal for assessing long-term effects, such research is still sparse for source credibility assessment interventions.

From an educational perspective, source credibility assessment interventions should aim for high ecological validity. While these learning environments are harder to design and control, they still seem beneficial to increase initial learning gain and the chance of participants transferring their new knowledge to their everyday lives for sustained effects. Research on the actual presence of such intended learning transfer would be an important next step. Hybrid learning interventions could be particularly effective, blending the flexibility and resources of digital tools with the personal connection of in-person instruction. This approach would allow educators to focus on guiding highly complex tasks, while workplace and adult learners can benefit from the flexibility of digital learning combined with face-to-face support.

12. Conclusion

While interventions on fostering source credibility assessment skills are overall moderately effective, we found that lateral reading interventions are more effective than the other source credibility assessment approaches. We interpret that the strategy is more closely related to learners' actual information environments and search behavior. Furthermore, source credibility assessment strategies in general are fairly complex, making such interventions a better fit for experienced learners and educational settings, which can feature longer curricular interventions. While the interventions are effective regardless of participant age and do not necessarily require in-person settings, digital technologies like the open Internet may still be beneficial, potentially increasing ecological validity. Future interventions should pay particular attention to long-term effectiveness, and try to further combine components from different source credibility assessment approaches.

CRedit authorship contribution statement

Marvin Fendt: Writing – review & editing, Writing – original draft, Methodology, Formal analysis, Data curation, Conceptualization. **Xenia Muth:** Writing – review & editing, Writing – original draft, Data curation. **Peter Adriaan Edelsbrunner:** Writing – review & editing, Writing – original draft, Supervision, Methodology.

Declaration of Generative AI and AI-assisted technologies in the writing process

During the preparation of this work, the authors used ChatGPT-4o and DeepL to improve the readability and language. After using these tools, the authors reviewed and edited the content as needed and take full responsibility for the content of the published article.

Declaration of competing interest

The authors have no conflict of interest.

Acknowledgments

We would like to thank the graduate students Selin Saygili and Ekin Arslan, who participated in the “Meta-Analysis” course at the Faculty of Psychology and Educational Sciences of the Ludwig-Maximilians-Universität München, Germany, for their invaluable contributions to the study coding.

References

- Abed, F., & Barzilai, S. (2023). Can students evaluate scientific YouTube videos? Examining students' strategies and criteria for evaluating videos versus webpages on climate change. *Journal of Computer Assisted Learning*, 39(2), 558–577. <http://dx.doi.org/10.1111/jcal.12762>.
- Aleven, V. A., & Koedinger, K. R. (2002). An effective metacognitive strategy: Learning by doing and explaining with a computer-based cognitive tutor. *Cognitive Science*, 26(2), 147–179. http://dx.doi.org/10.1207/s15516709cog2602_1.
- Alon, A. T., Rahimi, I. D., & Tahar, H. (2024). Fighting fake news on social media: A comparative evaluation of digital literacy interventions. *Current Psychology*, 43(19), 17343–17361. <http://dx.doi.org/10.1007/s12144-024-05668-4>.
- Anmarkrud, Ø., Bråten, I., Florit, E., & Mason, L. (2022). The role of individual differences in sourcing: A systematic review. *Educational Psychology Review*, 34(2), 749–792. <http://dx.doi.org/10.1007/s10648-021-09640-7>.
- Anmarkrud, Ø., Bråten, I., & Strømso, H. I. (2014). Multiple-documents literacy: Strategic processing, source awareness, and argumentation when reading multiple conflicting documents. *Learning and Individual Differences*, 30, 64–76. <http://dx.doi.org/10.1016/j.lindif.2013.01.007>.
- Argelagos, E., & Pifarre, M. (2012). Improving information problem solving skills in secondary education through embedded instruction. *Computers in Human Behavior*, 28(2), 515–526. <http://dx.doi.org/10.1016/j.chb.2011.10.024>.
- Axelsson, C.-A. W., & Nygren, T. (2024). The advantage of videos over text to boost adolescents' lateral reading in a digital workshop. *Behaviour & Information Technology*, 1–15. <http://dx.doi.org/10.1080/0144929X.2024.2308046>.
- Bartoš, F., & Schimmack, U. (2022). Z-curve 2.0: estimating replication rates and discovery rates. *Meta-Psychology*, 6, <http://dx.doi.org/10.15626/MP.2021.2720>.
- Barzilai, S., & Ka'adan, I. (2017). Learning to integrate divergent information sources: The interplay of epistemic cognition and epistemic metacognition. *Metacognition and Learning*, 12(2), 193–232. <http://dx.doi.org/10.1007/s11409-016-9165-7>.
- Barzilai, S., Mor-Hagani, S., Abed, F., Tal-Savir, D., Goldik, N., Talmon, I., et al. (2023). Misinformation is contagious: Middle school students learn how to evaluate and share information responsibly through a digital game. *Computers & Education*, 202, Article 104832. <http://dx.doi.org/10.1016/j.compedu.2023.104832>.
- Barzilai, S., Mor-Hagani, S., Zohar, A. R., Shlomi-Elou, T., & Ben-Yishai, R. (2020). Making sources visible: Promoting multiple document literacy with digital epistemic scaffolds. *Computers & Education*, 157, <http://dx.doi.org/10.1016/j.compedu.2020.103980>.
- Barzilai, S., & Stadler, M. (2025). Learning to evaluate (mis)information in an online game: strategies matter!. *Computers & Education*, 227, Article 105210. <http://dx.doi.org/10.1016/j.compedu.2024.105210>.
- Barzilai, S., Tal-Savir, D., Abed, F., Mor-Hagani, S., & Chinn, C. A. (2025). Scaffolding source evaluation during document-based scientific inquiry: The contributions of document mapping and shared criteria scaffolds. *Computers in Human Behavior*, 165, Article 108547. <http://dx.doi.org/10.1016/j.chb.2024.108547>.
- Barzilai, S., Thomm, E., & Shlomi-Elou, T. (2020). Dealing with disagreement: The roles of topic familiarity and disagreement explanation in evaluation of conflicting expert claims and sources. *Learning and Instruction*, 69, Article 101367. <http://dx.doi.org/10.1016/j.learninstruc.2020.101367>.
- Blatchford, P., Bassett, P., & Brown, P. (2011). Examining the effect of class size on classroom engagement and teacher–pupil interaction: differences in relation to pupil prior attainment and primary vs. Secondary schools. *Learning and Instruction*, 21(6), 715–730. <http://dx.doi.org/10.1016/j.learninstruc.2011.04.001>.
- Borenstein, M., Hedges, L. V., Higgins, J. P. T., & Rothstein, H. R. (2009). *Introduction to meta-analysis* (1st ed.). Wiley, <http://dx.doi.org/10.1002/9780470743386>.
- Braasch, J. L., Bråten, I., Strømso, H. I., Anmarkrud, Ø., & Ferguson, L. E. (2013). Promoting secondary school students' evaluation of source features of multiple documents. *Contemporary Educational Psychology*, 38(3), 180–195. <http://dx.doi.org/10.1016/j.cedpsych.2013.03.003>.
- Braasch, J. L. G., McCabe, R. M., & Daniel, F. (2016). Content integration across multiple documents reduces memory for sources. *Reading and Writing*, 29(8), 1571–1598. <http://dx.doi.org/10.1007/s11145-015-9609-5>.
- Brante, E. W., & Strømso, H. I. (2018). Sourcing in text comprehension: A review of interventions targeting sourcing skills. *Educational Psychology Review*, 30(3), 773–799. <http://dx.doi.org/10.1007/s10648-017-9421-7>.
- Bråten, I., Anmarkrud, Ø., Brandmo, C., & Strømso, H. I. (2014). Developing and testing a model of direct and indirect relationships between individual differences, processing, and multiple-text comprehension. *Learning and Instruction*, 30, 9–24. <http://dx.doi.org/10.1016/j.learninstruc.2013.11.002>.

- Bråten, I., Brandmo, C., Ferguson, L. E., & Strømsø, H. I. (2022). Epistemic justification in multiple document literacy: A refutation text intervention. *Contemporary Educational Psychology*, 71, Article 102122. <http://dx.doi.org/10.1016/j.cedpsych.2022.102122>.
- Bråten, I., Brante, E. W., & Strømsø, H. I. (2018). What really matters: The role of behavioural engagement in multiple document literacy tasks. *Journal of Research in Reading*, 41(4), 680–699. <http://dx.doi.org/10.1111/1467-9817.12247>.
- Bråten, I., Brante, E. W., & Strømsø, H. I. (2019). Teaching sourcing in upper secondary school: A comprehensive sourcing intervention with follow-up data. *Reading Research Quarterly*, 54(4), 481–505. <http://dx.doi.org/10.1002/rrq.253>.
- Bråten, I., Stadler, M., & Salmerón, L. (2018). The role of sourcing in discourse comprehension. In M. F. Schober, D. N. Rapp, & M. A. Britt (Eds.), *Routledge handbooks in linguistics, The routledge handbook of discourse processes* (2nd ed.). (pp. 141–166). New York, NY: Routledge/Taylor & Francis Group, <http://dx.doi.org/10.4324/9781315687384-10>.
- Breakstone, J., Smith, M., Wineburg, S., Rapaport, A., Carle, J., Garland, M., et al. (2021). Students' civic online reasoning: A national portrait. *Educational Researcher*, 50(8), 505–515. <http://dx.doi.org/10.3102/0013189X211017495>.
- Britt, M. A., & Aglinskas, C. (2002). Improving students' ability to identify and use source information. *Cognition and Instruction*, 20(4), 485–522. http://dx.doi.org/10.1207/S1532690XCI2004_2.
- Britt, M. A., Perfetti, C. A., Sandak, R., & Rouet, J.-F. (1999). Content integration and source separation in learning from multiple texts. In *Narrative comprehension, causality, and coherence: essays in honor of tom trabasso* (pp. 209–233). Mahwah, NJ, US: Lawrence Erlbaum Associates Publishers.
- Britt, M. A., Richter, T., & Rouet, J.-F. (2014). Scientific literacy: The role of goal-directed reading and evaluation in understanding scientific information. *Educational Psychologist*, 49(2), 104–122. <http://dx.doi.org/10.1080/00461520.2014.916217>.
- Britt, M. A., & Rouet, J.-F. (2012). Learning with multiple documents: component skills and their acquisition. In *Enhancing the quality of learning: dispositions, instruction, and learning processes* (pp. 276–314). New York, NY, US: Cambridge University Press, <http://dx.doi.org/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* (1). (pp. 17–33). Routledge, <http://dx.doi.org/10.4324/9781315627496-2>.
- Brodsky, J. E. (2022). *Fostering college students' fact-checking skills: three studies assessing lateral reading instruction in a general education course* (Ph.D. thesis), City University of Washington.
- Brodsky, J. E., Brooks, P. J., Pavlounis, D., & Johnston, J. L. (2023). Instruction increases Canadian students' preference for and use of lateral reading strategies to fact-check online information. *AERA Open*, 9, <http://dx.doi.org/10.1177/23328584231192106>.
- Brodsky, J. E., Brooks, P. J., Scimeca, D., Todorova, R., Galati, P., Batson, M., et al. (2021). Improving college students' fact-checking strategies through lateral reading instruction in a general education civics course. *Cognitive Research: Principles and Implications*, 6(1), 23. <http://dx.doi.org/10.1186/s41235-021-00291-4>.
- Brooks, D. C. (2011). Space matters: The impact of formal learning environments on student learning. *British Journal of Educational Technology*, 42(5), 719–726. <http://dx.doi.org/10.1111/j.1467-8535.2010.01098.x>.
- Burnett, L., & Cuevas, J. (2023). Using historical thinking strategies for improving elementary students' content knowledge and attitudes towards social studies. *Georgia Educational Researcher*, 20(1), 28–96. <http://dx.doi.org/10.20429/ger.2023.200102>.
- Capewell, G., Maertens, R., Remshard, M., Van Der Linden, S., Compton, J., Lewandowsky, S., et al. (2024). Misinformation interventions decay rapidly without an immediate posttest. *Journal of Applied Social Psychology*, 54(8), 441–454. <http://dx.doi.org/10.1111/jasp.13049>.
- Carter, E. C., Schönbrodt, F. D., Gervais, W. M., & Hilgard, J. (2019). Correcting for bias in psychology: A comparison of meta-analytic methods. *Advances in Methods and Practices in Psychological Science*, 2(2), 115–144. <http://dx.doi.org/10.1177/2515245919847196>.
- Chaiken, S. (1987). The heuristic model of persuasion. In *Social influence* (pp. 3–39). Psychology Press, <http://dx.doi.org/10.4324/9781315802121-2>.
- Chen, S., & Chaiken, S. (1999). The heuristic-systematic model in its broader context. In *Dual-process theories in social psychology* (pp. 73–96). New York, NY, US: The Guilford Press.
- Chernikova, O., Heitzmann, N., Stadler, M., Holzberger, D., Seidel, T., & Fischer, F. (2020). Simulation-based learning in higher education: A meta-analysis. *Review of Educational Research*, 90(4), 499–541. <http://dx.doi.org/10.3102/0034654320933544>.
- Costello, T. H., Pennycook, G., & Rand, D. G. (2024). Durably reducing conspiracy beliefs through dialogues with AI. *Science*, 385(6714), eadq1814. <http://dx.doi.org/10.1126/science.adq1814>.
- De La Paz, S. (2005). Effects of historical reasoning instruction and writing strategy mastery in culturally and academically diverse middle school classrooms. *Journal of Educational Psychology*, 97(2), 139–156. <http://dx.doi.org/10.1037/0022-0663.97.2.139>.
- De La Paz, S., & Felton, M. K. (2010). Reading and writing from multiple source documents in history: effects of strategy instruction with low to average high school writers. *Contemporary Educational Psychology*, 35(3), 174–192. <http://dx.doi.org/10.1016/j.cedpsych.2010.03.001>.
- De La Paz, S., Felton, M., Monte-Sano, C., Croninger, R., Jackson, C., Deogracias, J. S., et al. (2014). Developing historical reading and writing with adolescent readers: Effects on student learning. *Theory & Research in Social Education*, 42(2), 228–274. <http://dx.doi.org/10.1080/00933104.2014.908754>.
- De La Paz, S., Monte-Sano, C., Felton, M., Croninger, R., Jackson, C., & Piantadosi, K. W. (2017). A historical writing apprenticeship for adolescents: integrating disciplinary learning with cognitive strategies. *Reading Research Quarterly*, 52(1), 31–52. <http://dx.doi.org/10.1002/rrq.147>.
- Degner, M., Moser, S., & Lewalter, D. (2022). Digital media in institutional informal learning places: A systematic literature review. *Computers and Education Open*, 3, Article 100068. <http://dx.doi.org/10.1016/j.cao.2021.100068>.
- Delgado, P., Stang Lund, E., Salmerón, L., & Bråten, I. (2020). To click or not to click: Investigating conflict detection and sourcing in a multiple document hypertext environment. *Reading and Writing*, 33(8), 2049–2072. <http://dx.doi.org/10.1007/s11145-020-10030-8>.
- Dyrvold, A., & Bergvall, I. (2023). Static, dynamic and interactive elements in digital teaching materials in mathematics: How do they foster interaction, exploration and persistence? *LUMAT: International Journal on Math, Science and Technology Education*, 11(3), <http://dx.doi.org/10.31129/LUMAT.11.3.1941>.
- Ecker, U. K. H., Lewandowsky, S., Cook, J., Schmid, P., Fazio, L. K., Brashier, N., et al. (2022). The psychological drivers of misinformation belief and its resistance to correction. *Nature Reviews Psychology*, 1(1), 13–29. <http://dx.doi.org/10.1038/s44159-021-00006-y>.
- Edelsbrunner, P. A., Simonsmeier, B. A., & Schneider, M. (2025). The cronbach's alpha of domain-specific knowledge tests before and after learning: A meta-analysis of published studies. *Educational Psychology Review*, 37(1), 4. <http://dx.doi.org/10.1007/s10648-024-09982-y>.
- Eisend, M. (2006). Source credibility dimensions in marketing communication—A generalized solution. *Journal of Empirical Generalisations in Marketing Science*, 10(2).
- Elstad, E. (2024). AI in education: Rationale, principles, and instructional implications. <http://dx.doi.org/10.48550/ARXIV.2412.12116>.
- Fendt, M., Hufendiek, P., Oberparleiter, P., Scheibenzuber, C., & Edelsbrunner, P. A. (2025). The relationship of design elements with training effectiveness in a lateral reading training based on cognitive apprenticeship. Manuscript Submitted for Publication.
- Fendt, M., Nistor, N., Scheibenzuber, C., & Artmann, B. (2023). Sourcing against misinformation: Effects of a scalable lateral reading training based on cognitive apprenticeship. *Computers in Human Behavior*, 146, Article 107820. <http://dx.doi.org/10.1016/j.chb.2023.107820>.
- Fendt, M., Scheibenzuber, C., Edelsbrunner, P. A., & Nistor, N. (2025). Read between the li(ne)s: pedagogical lateral reading training to enhance information discernment. Unpublished Manuscript.
- Finn, J. D., Gerber, S. B., & Boyd-Zaharias, J. (2005). Small classes in the early grades, academic achievement, and graduating from high school. *Journal of Educational Psychology*, 97(2), 214–223. <http://dx.doi.org/10.1037/0022-0663.97.2.214>.
- Flanagin, A. J., & Metzger, M. J. (2008). Digital media and youth: unparalleled opportunity and unprecedented responsibility. In M. J. Metzger, & A. J. Flanagin (Eds.), *Digital media, youth, and credibility* (pp. 5–27). Cambridge, MA: The MIT Press, <http://dx.doi.org/10.1162/dmal.9780262562324.005>.
- Forzani, E. (2020). A three-tiered framework for proactive critical evaluation during online inquiry. *Journal of Adolescent & Adult Literacy*, 63(4), 401–414. <http://dx.doi.org/10.1002/jaal.1004>.
- Geist, E. A., & King, M. (2008). Different, not better: gender differences in mathematics learning and achievement. *Journal of Instructional Psychology*.
- Goldberg, T., Schwarz, B. B., & Porat, D. (2011). "Could they do it differently?": narrative and argumentative changes in students' writing following discussion of "hot" historical issues. *Cognition and Instruction*, 29(2), 185–217. <http://dx.doi.org/10.1080/07370008.2011.556832>.
- González-Gómez, F., Guardiola, J., Martín Rodríguez, Ó., & Montero Alonso, M. Á. (2012). Gender differences in E-learning satisfaction. *Computers & Education*, 58(1), 283–290. <http://dx.doi.org/10.1016/j.compedu.2011.08.017>.
- Graves, L. (2016). *Deciding what's true: The rise of political fact-checking in American journalism*. New York: Columbia University Press.
- Griffin, T. D., Jaeger, A. J., Britt, M. A., & Wiley, J. (2024). Improving multiple document comprehension with a lesson about multi-causal explanations in science. *Instructional Science*, 52(4), 639–664. <http://dx.doi.org/10.1007/s11251-023-09657-1>.
- Gwet, K. L. (2014). *Handbook of inter-rater reliability: The definitive guide to measuring the extent of agreement among raters*. Advanced Analytics, LLC.
- Hämäläinen, E. K., Kiili, C., Marttunen, M., Räikkönen, E., González-Ibáñez, R., & Leppänen, P. H. (2020). Promoting sixth graders' credibility evaluation of web pages: An intervention study. *Computers in Human Behavior*, 110, Article 106372. <http://dx.doi.org/10.1016/j.chb.2020.106372>.
- Hämäläinen, E. K., Kiili, C., Räikkönen, E., Lakkala, M., Ilomäki, L., Toom, A., et al. (2023). Teaching sourcing during online inquiry – adolescents with the weakest skills benefited the most. *Instructional Science*, 51(1), 135–163. <http://dx.doi.org/10.1007/s11251-022-09597-2>.

- Hammer, D., & Berland, L. K. (2014). Confusing claims for data: A critique of common practices for presenting qualitative research on learning. *Journal of the Learning Sciences*, 23(1), 37–46. <http://dx.doi.org/10.1080/10508406.2013.802652>.
- Hoekstra, A., Beijard, D., Brekelmans, M., & Korthagen, F. (2007). Experienced teachers' informal learning from classroom teaching. *Teachers and Teaching*, 13(2), 189–206. <http://dx.doi.org/10.1080/13540600601152546>.
- Holmes, W., & Tuomi, I. (2022). State of the art and practice in AI in education. *European Journal of Education*, 57(4), 542–570. <http://dx.doi.org/10.1111/ejed.12533>.
- Hou, X., Nguyen, H. A., Richey, J. E., & McLaren, B. M. (2020). Exploring how gender and enjoyment impact learning in a digital learning game. In I. I. Bittencourt, M. Cukurova, K. Muldner, R. Luckin, & E. Millán (Eds.), *Artificial intelligence in education: vol. 12163*, (pp. 255–268). Cham: Springer International Publishing, http://dx.doi.org/10.1007/978-3-030-52237-7_21.
- Hunter, L. Y. (2023). Social media, disinformation, and democracy: How different types of social media usage affect democracy cross-nationally. *Democratization*, 30(6), 1040–1072. <http://dx.doi.org/10.1080/13510347.2023.2208355>.
- Hunter, J. E., & Schmidt, F. L. (2004). *Methods of meta-analysis: Correcting error and bias in research findings*. Sage.
- Imbens, G. W. (2024). Causal inference in the social sciences. *Annual Review of Statistics and its Application*, 11(1), 123–152. <http://dx.doi.org/10.1146/annurev-statistics-033121-114601>.
- Int'Hout, J., Ioannidis, J. P. A., Rovers, M. M., & Goeman, J. J. (2016). Plea for routinely presenting prediction intervals in meta-analysis. *BMJ Open*, 6(7), Article e010247. <http://dx.doi.org/10.1136/bmjopen-2015-010247>.
- Johnson, M., & Majewska, D. (2022). Formal, non-formal, and informal learning: What are they, and how can we research them? research report. *Cambridge University Press & Assessment*.
- Kammerer, Y., Amann, D. G., & Gerjets, P. (2015). When adults without university education search the internet for health information: The roles of internet-specific epistemic beliefs and a source evaluation intervention. *Computers in Human Behavior*, 48, 297–309. <http://dx.doi.org/10.1016/j.chb.2015.01.045>.
- Kammerer, Y., Meier, N., & Stahl, E. (2016). Fostering secondary-school students' intertext model formation when reading a set of websites: The effectiveness of source prompts. *Computers & Education*, 102, 52–64. <http://dx.doi.org/10.1016/j.compedu.2016.07.001>.
- Kenski, K., & Stroud, N. J. (2006). Connections between internet use and political efficacy, knowledge, and participation. *Journal of Broadcasting & Electronic Media*, 50(2), 173–192. http://dx.doi.org/10.1207/s15506878jebem5002_1.
- Kim, S. M., & Hannafin, M. J. (2016). Synergies: Effects of source representation and goal instructions on evidence quality, reasoning, and conceptual integration during argumentation-driven inquiry. *Instructional Science*, 44(5), 441–476. <http://dx.doi.org/10.1007/s11251-016-9381-1>.
- Kingsley, T. L., Cassidy, J. C., & Tancock, S. M. (2015). Successfully promoting 21st century online research skills: Interventions in 5th-grade classrooms. *Reading Horizons: A Journal of Literacy and Language Arts*, 54(2), 5.
- Kirschner, P. A., & van Merriënboer, J. J. (2013). Do learners really know best? urban legends in education. *Educational Psychologist*, 48(3), 169–183. <http://dx.doi.org/10.1080/00461520.2013.804395>.
- Kozyreva, A., Wineburg, S., Lewandowsky, S., & Hertwig, R. (2023). Critical ignoring as a core competence for digital citizens. *Current Directions in Psychological Science*, 32(1), 81–88. <http://dx.doi.org/10.1177/09637214221121570>.
- Kuiper, E., Volman, M., & Terwel, J. (2009). Developing web literacy in collaborative inquiry activities. *Computers & Education*, 52, 668–680. <http://dx.doi.org/10.1016/j.compedu.2008.11.010>.
- Lee, Y.-H. (2021). Scaffolding university students' epistemic cognition during multi-modal multiple-document reading: The effects of the epistemic prompting and the automated reflection report. *The Internet and Higher Education*, 49, Article 100777. <http://dx.doi.org/10.1016/j.iheduc.2020.100777>.
- Lee, Y.-H. (2022). Beyond online search strategies: The effects of internet epistemic beliefs and different note-taking formats on online multiple document reading comprehension. *Journal of Computer Assisted Learning*, 38(4), 1102–1114. <http://dx.doi.org/10.1111/jcal.12668>.
- Lee, A. Y., Moore, R. C., & Hancock, J. T. (2024). Building resilience to misinformation in communities of color: Results from two studies of tailored digital media literacy interventions. *New Media & Society*, Article 14614448241227841. <http://dx.doi.org/10.1177/14614448241227841>.
- Leeder, C., & Shah, C. (2016). Practicing critical evaluation of online sources improves student search behavior. *The Journal of Academic Librarianship*, 42(4), 459–468. <http://dx.doi.org/10.1016/j.acalib.2016.04.001>.
- Lescarret, C., Magnier, J., Le Floch, V., Sakdavong, J.-C., Boucheix, J.-M., & Amadiou, F. (2024). Do you trust this speaker? The impact of prompting on middle-school students' consideration of source when watching conflicting videos. *Instructional Science*, 52(1), 41–69. <http://dx.doi.org/10.1007/s11251-023-09637-5>.
- List, A., & Alexander, P. A. (2017). Cognitive affective engagement model of multiple source use. *Educational Psychologist*, 52(3), 182–199. <http://dx.doi.org/10.1080/00461520.2017.1329014>.
- List, A., Grossnickle, E. M., & Alexander, P. A. (2016). Undergraduate students' justifications for source selection in a digital academic context. *Journal of Educational Computing Research*, 54(1), 22–61. <http://dx.doi.org/10.1177/0735633115606659>.
- Macedo-Rouet, M., Braasch, J. L., Britt, M. A., & Rouet, J.-F. (2013). Teaching fourth and fifth graders to evaluate information sources during text comprehension. *Cognition and Instruction*, 31(2), 204–226. <http://dx.doi.org/10.1080/07370008.2013.769995>.
- Marten, P. L., Alßmann, S., Baumgarten-Kelm, C., & Stadler, M. (2025). Did 5G radiation really kill birds? Training lower secondary students in epistemic strategies to counter online misinformation. *Learning and Individual Differences*, 122, Article 102685. <http://dx.doi.org/10.1016/j.lindif.2025.102685>.
- Marten, P. L., & Stadler, M. (2025). Building resilience against online misinformation: A teacher-led training promoting evaluation strategies among lower secondary students. *Computers in Human Behavior*, 165, Article 108548. <http://dx.doi.org/10.1016/j.chb.2024.108548>.
- Martin, F., Sun, T., & Westine, C. D. (2020). A systematic review of research on online teaching and learning from 2009 to 2018. *Computers & Education*, 159, Article 104009. <http://dx.doi.org/10.1016/j.compedu.2020.104009>.
- Martínez, M. A., Saux, G., Londra, F., & Burin, D. I. (2024). Effects of a classroom intervention on college students' sourcing skills: Replication and extension study. *Discourse Processes*, 61(4–5), 255–280. <http://dx.doi.org/10.1080/0163853X.2024.2339738>.
- Martini, C., Floris, M., Ronzani, P., Ausili, L., Adorno, G., Pennacchioni, G., et al. (2025). The impact of interventions against science disinformation in high school students. *Scientific Reports*, <http://dx.doi.org/10.17605/osc.io/QKPB5>.
- Mason, L., Junyent, A. A., & Tornatora, M. C. (2014). Epistemic evaluation and comprehension of web-source information on controversial science-related topics: Effects of a short-term instructional intervention. *Computers & Education*, 76, 143–157. <http://dx.doi.org/10.1016/j.compedu.2014.03.016>.
- Mason, L., Moè, A., Tornatora, M. C., & Ronconi, A. (2022). Promoting web-source evaluation and comprehension of conflicting online documents: effects of classroom interventions. In P. Limone, R. Di Fuccio, & G. A. Toto (Eds.), *Psychology, learning, technology: vol. 1606*, (pp. 3–21). Cham: Springer International Publishing.
- McBride, R. E., Xiang, P., Wittenburg, D., & Shen, J. (2002). An analysis of preservice teachers' dispositions toward critical thinking: A cross-cultural perspective. *Asia-Pacific Journal of Teacher Education*, 30(2), 131–140. <http://dx.doi.org/10.1080/13598660220135649>.
- McCroskey, J. C. (1966). Scales for the measurement of ethos. *Speech Monographs*, 33(1), 65–72. <http://dx.doi.org/10.1080/03637756609375482>.
- McGrew, S. (2024). Teaching lateral reading: Interventions to help people read like fact checkers. *Current Opinion in Psychology*, 55, Article 101737. <http://dx.doi.org/10.1016/j.copsyc.2023.101737>.
- McGrew, S., & Byrne, V. L. (2020). Who is behind this? Preparing high school students to evaluate online content. *Journal of Research on Technology in Education*, 53(4), 457–475. <http://dx.doi.org/10.1080/15391523.2020.1795956>.
- McGrew, S., Smith, M., Breakstone, J., Ortega, T., & Wineburg, S. (2019). Improving university students' web savvy: an intervention study. *British Journal of Educational Psychology*, 89(3), 485–500. <http://dx.doi.org/10.1111/bjep.12279>.
- Meier, D. J. (2010). *The use of primary source historical documents, historical reasoning heuristics, and the subsequent development of historical empathy* (Ph.D. thesis), Temple University, <http://dx.doi.org/10.34944/dspace/1880>.
- Metzger, M. J. (2007). Making sense of credibility on the web: models for evaluating online information and recommendations for future research. *Journal of the American Society for Information Science and Technology*, 58(13), 2078–2091. <http://dx.doi.org/10.1002/asi.20672>.
- Metzger, M. J., Flanagin, A. J., & Medders, R. B. (2010). Social and heuristic approaches to credibility evaluation online. *Journal of Communication*, 60(3), 413–439. <http://dx.doi.org/10.1111/j.1460-2466.2010.01488.x>.
- Moore, R. C., & Hancock, J. T. (2022). A digital media literacy intervention for older adults improves resilience to fake news. *Scientific Reports*, 12(1), 6008. <http://dx.doi.org/10.1038/s41598-022-08437-0>.
- Muehlhoff, R., & Henningsen, M. (2024). Chatbots im schulunterricht: Wir testen das fobizz-tool zur automatischen bewertung von hausaufgaben. <http://dx.doi.org/10.48550/ARXIV.2412.06651>.
- Muis, K. R., Denton, C., & Dubé, A. (2022). Identifying CRAAP on the internet: A source evaluation intervention. *Advances in Social Sciences Research Journal*, 9(7), 239–265. <http://dx.doi.org/10.14738/assrj.97.12670>.
- Nielsen, T., Martínez-García, I., & Alastor, E. (2021). Critical thinking of psychology students: A within- and cross-cultural study using Rasch models. *Scandinavian Journal of Psychology*, 62(3), 426–435. <http://dx.doi.org/10.1111/sjop.12714>.
- Niklas, F., Birtwistle, E., Wirth, A., Schiele, T., & Mues, A. (2022). App-based learning for kindergarten children at home (Learning4Kids): Study protocol for cohort 2 and the school assessments. *BMC Pediatrics*, 22(1), 705. <http://dx.doi.org/10.1186/s12887-022-03737-w>.
- Nokes, J. D. (2014). Elementary students' roles and epistemic stances during document-based history lessons. *Theory and Research in Social Education*, 42(3), 375–413. <http://dx.doi.org/10.1080/00933104.2014.937546>.
- Nokes, J. D., Dole, J. A., & Hacker, D. J. (2007). Teaching high school students to use heuristics while reading historical texts. *Journal of Educational Psychology*, 99(3), 492–504. <http://dx.doi.org/10.1037/0022-0663.99.3.492>.
- Ohanian, R. (1990). Construction and validation of a scale to measure celebrity endorsers' perceived expertise, trustworthiness, and attractiveness. *Journal of Advertising*, 19(3), 39–52. <http://dx.doi.org/10.1080/00913367.1990.10673191>.

- Osborne, J., & Pimentel, D. (2022). Science, misinformation, and the role of education. *Science*, 378(6617), 246–248. <http://dx.doi.org/10.1126/science.abq8093>.
- Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., et al. (2021). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *BMJ*, n71. <http://dx.doi.org/10.1136/bmj.n71>.
- Panizza, F., Ronzani, P., Martini, C., Mattavelli, S., Morisseau, T., & Motterlini, M. (2022). Lateral reading and monetary incentives to spot disinformation about science. *Scientific Reports*, 12(1), 5678. <http://dx.doi.org/10.1038/s41598-022-09168-y>.
- Patzl, S., Oberleiter, S., & Pietschnig, J. (2024). Mirror, mirror on the wall: A meta-analysis on the validity of self-assessed intelligence through the lens of the multiverse. *Journal of Intelligence*, 12(9), 81. <http://dx.doi.org/10.3390/jintelligence12090081>.
- Paul, J., Stadler, M., & Bromme, R. (2019). Effects of a sourcing prompt and conflicts in reading materials on elementary students' use of source information. *Discourse Processes*, 56(2), 155–169. <http://dx.doi.org/10.1080/0163853X.2017.1402165>.
- Peng, D., & Yu, Z. (2022). A literature review of digital literacy over two decades. In E. Rezvani (Ed.), *Education Research International*, 2022, 1–8. <http://dx.doi.org/10.1155/2022/2533413>.
- Pennycook, G., & Rand, D. G. (2019). Lazy, not biased: susceptibility to partisan fake news is better explained by lack of reasoning than by motivated reasoning. *Cognition*, 188, 39–50. <http://dx.doi.org/10.1016/j.cognition.2018.06.011>.
- Pennycook, G., & Rand, D. G. (2022). Accuracy prompts are a replicable and generalizable approach for reducing the spread of misinformation. *Nature Communications*, 13(1), 2333. <http://dx.doi.org/10.1038/s41467-022-30073-5>.
- Pérez, A., Potocki, A., Stadler, M., Macedo-Rouet, M., Paul, J., Salmerón, L., et al. (2018). Fostering teenagers' assessment of information reliability: effects of a classroom intervention focused on critical source dimensions. *Learning and Instruction*, 58, 53–64. <http://dx.doi.org/10.1016/j.learninstruc.2018.04.006>.
- Perfetti, C. A., Rouet, J.-F., & Britt, M. A. (1999). Toward a theory of documents representation. In *The construction of mental representations during reading* (pp. 99–122). Mahwah, NJ, US: Lawrence Erlbaum Associates Publishers.
- Petty, R. E., & Cacioppo, J. T. (1986). The elaboration likelihood model of persuasion. In L. Berkowitz (Ed.), *Advances in experimental social psychology*: vol. 19, (pp. 123–205). Academic Press, [http://dx.doi.org/10.1016/S0065-2601\(08\)60214-2](http://dx.doi.org/10.1016/S0065-2601(08)60214-2).
- Petty, R. E., & Wegener, D. T. (1999). The elaboration likelihood model: current status and controversies. In *Dual-process theories in social psychology* (pp. 37–72). New York, NY, US: The Guilford Press.
- Pornpitakpan, C. (2004). The persuasiveness of source credibility: A critical review of five decades' evidence. *Journal of Applied Social Psychology*, 34(2), 243–281. <http://dx.doi.org/10.1111/j.1559-1816.2004.tb02547.x>.
- Ranger, J., Kuhn, J. T., & Ortner, T. M. (2020). Modeling responses and response times in tests with the hierarchical model and the three-parameter lognormal distribution. *Educational and Psychological Measurement*, 80(6), 1059–1089.
- Raudenbush, S. W., & Bryk, A. S. (2002). Hierarchical linear models: Applications and data analysis methods, In *Advanced quantitative techniques in the social sciences*, (2nd ed.). (1), Thousand Oaks: Sage Publications.
- Reisman, A. (2012). Reading like a historian: A document-based history curriculum intervention in urban high schools. *Cognition and Instruction*, 30(1), 86–112. <http://dx.doi.org/10.1080/07370008.2011.634081>.
- Rouet, J.-F., & Britt, M. A. (2011). Relevance processes in multiple document comprehension. In *Text relevance and learning from text* (pp. 19–52). Charlotte, NC, US: IAP Information Age Publishing.
- Rubin, D. B. (1976). Inference and missing data. *Biometrika*, 63(3), 581–592. <http://dx.doi.org/10.1093/biomet/63.3.581>.
- Scharrer, L., & Salmerón, L. (2016). Sourcing in the reading process: Introduction to the special issue. *Reading and Writing*, 29(8), 1539–1548. <http://dx.doi.org/10.1007/s11145-016-9676-2>.
- Singer, J. B. (2023). Closing the barn door? fact-checkers as retroactive gatekeepers of the COVID-19 “infodemic”. *Journalism & Mass Communication Quarterly*, 100(2), 332–353. <http://dx.doi.org/10.1177/10776990231168599>.
- Sonia, A. N., Magliano, J. P., McCarthy, K. S., Creer, S. D., McNamara, D. S., & Allen, L. (2022). Integration in multiple-document comprehension: A natural language processing approach. *Discourse Processes*, 59(5–6), 417–438. <http://dx.doi.org/10.1080/0163853X.2022.2079320>.
- Stadler, M., & Bromme, R. (2007). Dealing with multiple documents on the WWW: The role of metacognition in the formation of documents models. *International Journal of Computer-Supported Collaborative Learning*, 2(2–3), 191–210. <http://dx.doi.org/10.1007/s11412-007-9015-3>.
- Stadler, M., & Bromme, R. (2008). Effects of the metacognitive computer-tool met.a.ware on the web search of laypersons. In *Instructional support for enhancing students' information problem solving ability: Computers in Human Behavior*, In *Instructional support for enhancing students' information problem solving ability*: 24(3), 716–737. <http://dx.doi.org/10.1016/j.chb.2007.01.023>.
- Stadler, M., & Bromme, R. (2014). The content-source integration model: A taxonomic description of how readers comprehend conflicting scientific information. In *Processing inaccurate information: theoretical and applied perspectives from cognitive science and the educational sciences* (pp. 379–402). Cambridge, MA, US: The MIT Press, <http://dx.doi.org/10.7551/mitpress/9737.001.0001>.
- Stadler, M., Scharrer, L., Macedo-Rouet, M., Rouet, J.-F., & Bromme, R. (2016). Improving vocational students' consideration of source information when deciding about science controversies. *Reading and Writing*, 29(4), 705–729. <http://dx.doi.org/10.1007/s11145-016-9623-2>.
- Stanley, T. D., & Doucouliagos, H. (2014). Meta-regression approximations to reduce publication selection bias: T. d. Stanley and h. doucouliagos. *Research Synthesis Methods*, 5(1), 60–78. <http://dx.doi.org/10.1002/jrsm.1095>.
- Strømso, 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. <http://dx.doi.org/10.1080/07370008.2013.769994>.
- Sultan, M., Tump, A. N., Ehmann, N., Lorenz-Spreen, P., Hertwig, R., Gollwitzer, A., et al. (2024). Susceptibility to online misinformation: A systematic meta-analysis of demographic and psychological factors. *Proceedings of the National Academy of Sciences*, 121(47), Article e2409329121. <http://dx.doi.org/10.1073/pnas.2409329121>.
- Sundar, S. S. (2008). The MAIN model: A heuristic approach to understanding technology effects on credibility. In M. J. Metzger, & A. J. Flanagin (Eds.), *The john d. and catherine t. macArthur foundation series on digital media and learning*, *Digital Media, Youth, and Credibility, The john d. and catherine t. macArthur foundation series on digital media and learning*, 73–100. <http://dx.doi.org/10.1162/dmal.9780262562324.073>.
- Tang, C. M., & Chaw, L. Y. (2016). Digital literacy: A prerequisite for effective learning in a blended learning environment? *Electronic Journal of E-Learning*, 14(1), 54–65.
- Tipton, E., Bryan, C., Murray, J., McDaniel, M. A., Schneider, B., & Yeager, D. S. (2023). Why meta-analyses of growth mindset and other interventions should follow best practices for examining heterogeneity: commentary on Macnamara and Burgoyne (2023) and Burnette et al. (2023). *Psychological Bulletin*, 149(3–4), 229–241. <http://dx.doi.org/10.1037/bul0000384>.
- Tomasik, M. J., Helbling, L. A., & Moser, U. (2021). Educational gains of in-person vs. Distance learning in primary and secondary schools: A natural experiment during the COVID -19 pandemic school closures in Switzerland. *International Journal of Psychology*, 56(4), 566–576. <http://dx.doi.org/10.1002/ijop.12728>.
- Ulyshen, T. Z., Koehler, M. J., & Gao, F. (2015). Understanding the connection between epistemic beliefs and internet searching. *Journal of Educational Computing Research*, 53(3), 345–383. <http://dx.doi.org/10.1177/0735633115599604>.
- VanSledright, B. A. (2002). Fifth graders investigating history in the classroom: Results from a researcher-practitioner design experiment. *The Elementary School Journal*, 103(2), 131–160. <http://dx.doi.org/10.1086/499720>.
- Viechtbauer, W. (2010). Conducting meta-analyses in R with the metafor package. *Journal of Statistical Software*, 36(3), <http://dx.doi.org/10.18637/jss.v036.i03>.
- Viechtbauer, W., & Cheung, M. W.-L. (2010). Outlier and influence diagnostics for meta-analysis. *Research Synthesis Methods*, 1(2), 112–125. <http://dx.doi.org/10.1002/jrsm.11>.
- Walraven, A., Brand-Gruwel, S., & Boshuizen, H. P. A. (2013). Fostering students' evaluation behaviour while searching the internet. *Instructional Science*, 41(1), 125–146. <http://dx.doi.org/10.1007/s11251-012-9221-x>.
- Walter, N., Cohen, J., Holbert, R. L., & Morag, Y. (2020). Fact-checking: A meta-analysis of what works and for whom. *Political Communication*, 37(3), 350–375. <http://dx.doi.org/10.1080/10584609.2019.1668894>.
- Wigfield, A., & Eccles, J. S. (2000). Expectancy-value theory of achievement motivation. *Contemporary Educational Psychology*, 25(1), 68–81. <http://dx.doi.org/10.1006/ceps.1999.1015>.
- Wiley, J., Goldman, S. R., Graesser, A. C., Sanchez, C. A., Ash, I. K., & Hemmerich, J. A. (2009). Source evaluation, comprehension, and learning in internet science inquiry tasks. *American Educational Research Journal*, 46(4), 1060–1106. <http://dx.doi.org/10.3102/0002831209333183>.
- Wineburg, S. (1991). Historical problem solving: A study of the cognitive processes used in the evaluation of documentary and pictorial evidence. *Journal of Educational Psychology*, 83(1), 73–87. <http://dx.doi.org/10.1037/0022-0663.83.1.73>.
- Wineburg, S. (2024). The art of critical ignoring. *The Science Teacher*, 91(5), 16–20. <http://dx.doi.org/10.1080/00368555.2024.2388450>.
- Wineburg, S., Breakstone, J., McGrew, S., Smith, M. D., & 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. <http://dx.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: The Voice of Scholarship in Education*, 121(11), 1–40. <http://dx.doi.org/10.1177/016146811912101102>.
- Wissinger, D. R., De La Paz, S., & Jackson, C. (2021). The effects of historical reading and writing strategy instruction with fourth- through sixth-grade students. *Journal of Educational Psychology*, 113(1), 49–67. <http://dx.doi.org/10.1037/edu0000463>.
- Zhang, S., & Duke, N. K. (2011). The impact of instruction in the WWWDOT framework on students' disposition and ability to evaluate web sites as sources of information. *The Elementary School Journal*, 112(1), 132–154. <http://dx.doi.org/10.1086/660687>.
- Zhang, D., & Pradeep, R. (2023). ReadProbe: A demo of retrieval-enhanced large language models to support lateral reading. <http://dx.doi.org/10.48550/ARXIV.2306.07875>.
- Zwaan, R. A. (1994). Effect of genre expectations on text comprehension. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, 20(4), 920–933. <http://dx.doi.org/10.1037/0278-7393.20.4.920>.