

Students' abilities to evaluate the credibility of online texts: The role of internet-specific epistemic justifications

Elina K. Hämäläinen¹  | Carita Kiili^{2,3}  | Eija Räikkönen⁴  | Miika Marttunen¹ 

¹Department of Education, University of Jyväskylä, Jyväskylä, Finland

²Faculty of Education and Culture, Tampere University, Tampere, Finland

³Department of Education, University of Oslo, Norway

⁴Faculty of Education and Psychology, University of Jyväskylä, Jyväskylä, Finland

Correspondence

Elina K. Hämäläinen, University of Jyväskylä, P.O. Box 35, Jyväskylä 40014, Finland.
Email: elina.k.hamalainen@jyu.fi

Abstract

Previous evaluation studies have rarely used authentic online texts and investigated upper secondary school students' use of evaluation criteria and deep reasoning. The associations between internet-specific epistemic justifications for knowing and credibility evaluation of online texts are not yet fully understood among adolescents. This study investigated upper secondary school students' ($N = 372$) abilities to evaluate self-selected authentic online texts and the role of internet-specific epistemic justifications in students' evaluation performance when solving a health-related information problem. Students selected three texts with Google Custom Search Engine and evaluated their credibility. Students' evaluation performance across the three texts was determined according to the different aspects evaluated (author, venue, intentions, evidence and corroboration) and the depth of their evaluations. Students also filled in the Internet-Specific Epistemic Justifications (ISEJ) inventory previously validated with pre-service teachers. The results revealed considerable differences in students' abilities to evaluate online texts. Students' beliefs in justification by authority and justification by multiple sources positively predicted their evaluation performance similarly in both topics. The findings suggest that the ISEJ inventory is also valid for upper secondary school students. Students should be explicitly taught to evaluate different credibility aspects and scaffolded to deeply engage with online information.

KEYWORDS

adolescents, credibility evaluation, internet-specific epistemic justifications, justifications for knowing, online inquiry, sourcing

1 | INTRODUCTION

The current COVID-19 pandemic has challenged publics' abilities to evaluate the credibility of health information online. Misleading information has spread rapidly via the Internet. Moreover, experts may disagree in a novel uncertain situation where it takes time for scientific research to yield results. As a whole, the current online debate reflects a post-truth world in which laypersons may disagree about evidence-

based facts and place more weight on their personal beliefs than on scientific knowledge when deciding what to believe (Sinatra & Lombardi, 2020).

Aside from the pandemic, people's trust in inaccurate health information, or distrust of credible health information, can negatively influence their health and use of health care system resources (Freeman et al., 2020). A recent review (Freeman et al., 2020) showed that, for many adolescents, evaluating the credibility of health-related

This is an open access article under the terms of the Creative Commons Attribution-NonCommercial-NoDerivs License, which permits use and distribution in any medium, provided the original work is properly cited, the use is non-commercial and no modifications or adaptations are made.

© 2021 The Authors. *Journal of Computer Assisted Learning* published by John Wiley & Sons Ltd.

online information is challenging. However, even if adolescents seem to understand that online information is not always to be trusted, many remain unsure of how to evaluate its credibility (e.g., Freeman et al., 2020; Paul et al., 2017).

This study investigates upper secondary students' abilities to evaluate the credibility of self-selected health-related authentic online texts and their beliefs in justifications for knowing on the Internet, that is, the extent to which they rely on their prior knowledge, the expertise of the source and multiple online texts when judging the information they encounter online (Bråten et al., 2019). To further knowledge of adolescents' evaluation of online information, this study examines how students' beliefs in their justifications for knowing on the Internet were associated with their evaluation performance.

1.1 | Theoretical frameworks

In this study, we rely on two theoretical frameworks: online research and comprehension (Leu et al., 2019) and multiple documents comprehension (Perfetti et al., 1999). The model of online research and comprehension guided our construction of the online inquiry task while the theory of multiple documents comprehension formed the foundation for our analysis of students' credibility evaluations.

According to Leu et al. (2019), online research and comprehension or online inquiry is a process that requires online readers to make decisions about what to read, how to read and how to utilize texts to solve a problem. Online research comprises five cyclic processes: (1) asking questions and defining information need, (2) locating information with a search engine, (3) evaluating information, (4) synthesizing information and (5) communicating results to others. Ideally, readers evaluate texts during different phases of online inquiry (Gerjets et al., 2011; Rieh, 2002). First, when reading the search engine results page, readers have an opportunity to make predictive judgements to inform their selection of useful texts by utilizing title, URL address or example text (e.g., Rieh, 2002). However, readers tend to select links that are at the top of the search results (Gerjets et al., 2011; Pan et al., 2007). Second, evaluative judgements can take place after accessing the online text. When the evaluative judgement meets the predictive judgement, the reader decides to use the information or to stay on the page (Rieh, 2002). It has been shown that skilful readers make predictive and evaluative judgements continuously as an iterative process until they complete their searches (e.g., Rieh, 2002). Finally, skilful readers also compare and verify the information by evaluating the collection of selected texts (Gerjets et al., 2011; Meola, 2004).

The theory of multiple documents comprehension (Britt et al., 2018; Perfetti et al., 1999; Rouet, 2006) describes how, to achieve their reading goals, readers select, evaluate and use information from more than one document. Compared to single document comprehension, during which readers integrate text contents with their prior knowledge, multiple document comprehension presents additional challenges in building a coherent representation of the information contained in different documents. For example, contradictory information gathered from different sources might be difficult to

integrate coherently. To address these challenges, the documents model framework proposes that readers need to form two representations: an integrated mental model and an intertext model. The integrated mental model refers to the representation of contents across the documents organized in accordance with the reading task. The intertext model, in turn, refers to the representation of source information (e.g., authors' credentials and intentions) and links between the sources to its content and rhetorical relationships between the sources. By combining these models, readers can understand complex and potentially conflicting information by incorporating the contents of documents into their respective sources.

While sourcing (i.e., attending to, evaluating and using available information about the documents' source features) is a fundamental component in multiple document comprehension, it has recently received much attention among reading researchers (e.g., Brante & Strømsø, 2018; Bråten et al., 2018). The open nature of the Internet, where almost anyone can publish their views, has accelerated the need to understand the role of sourcing when readers engage in online inquiry. The next section discusses the essential source features in more detail.

1.2 | Evaluation of credibility

Because of the ease of publishing on the Internet and the absence of traditional gatekeepers, the Internet is a marketplace of opinions that can be presented by authors with different levels of knowledge (Salmerón et al., 2018). It is therefore essential to evaluate authors' expertise by paying attention to their credentials, affiliations and positions (e.g., Bråten et al., 2018). It is also worthwhile to consider the publication practices of the venue, that is, who is allowed to write the texts that constitutes a website and how the accuracy of information is ensured (Braasch et al., 2013).

Aside from their expertise, the authors' intention is the source feature considered to most merit critical evaluation (Bråten et al., 2018; Potocki et al., 2020). Readers can evaluate the intentions of authors or venues by considering the motives or interests behind the message. Is the author's purpose to share research-based knowledge, sell a product, or persuade? For example, recognizing commercial intentions seems to be difficult, particularly for adolescent readers (Kiili et al., 2018). Furthermore, research suggests that students tend to pay more attention to text content than to source features when evaluating online texts (e.g., Bråten, McCrudden, et al., 2018; Kiili et al., 2019).

Attending to source features provides useful cues for evaluating the evidence that authors rely on, especially when readers do not have much prior knowledge on the topic (Bråten, McCrudden, et al., 2018). It can reasonably be assumed that academics mostly base their arguments on research evidence whereas laypersons may rely more on personal experience (Hoeken, 2001). Besides, readers can evaluate the quality of the information sources (e.g., references cited, persons interviewed) that authors employ and how well the evidence given supports the claim (Sinatra & Lombardi, 2020). A recent study by Hämäläinen et al. (2020) showed that evaluating the evidence presented in online texts was challenging for adolescents.

Studies that have examined reading practices of experts (e.g., academic librarians, journalists or historians) have highlighted the importance of corroboration, that is, checking the accuracy of facts or statements from another information resource before accepting them as plausible (Kohnen & Mertens, 2019; Wineburg, 1991). The more online texts students encounter and compare the better they will become at assessing what counts as high-quality information and what does not (Meola, 2004). It is essential that corroboration is performed in relation to other credible documents instead of students' own prior knowledge and beliefs, as these may be biased (Greene et al., 2019; Sinatra & Lombardi, 2020).

In general, the various aspects of credibility are often intertwined. For example, online texts display rhetorical relations such as supporting (evidence and corroboration) and opposing (disagree, contradict) each other (Britt et al., 2018). Accordingly, conflicting information has been found to promote the evaluation and comparison of the sources of documents among older students (e.g., Kammerer et al., 2016; Rouet et al., 2016). In the present study, we used the above-introduced aspects of credibility: the author's expertise, venue, intentions, evidence and corroboration to assess students' performance in a credibility evaluation task.

1.3 | Justifications for knowing

The vast amount of easily accessible information and lack of traditional gatekeepers on the Internet set high demands on readers' epistemic cognition, that is, their abilities to construct, evaluate and use knowledge (Greene & Yu, 2015). More specifically, epistemic cognition comprises both epistemic beliefs and the application of those beliefs (e.g., Greene et al., 2008). Hofer and Pintrich (1997) presented four dimensions of epistemic beliefs about knowledge and knowing: (1) certainty of knowledge, (2) simplicity of knowledge, (3) source of knowledge and (4) justification for knowing. Epistemic beliefs, particularly justifications for knowing, can be applied, for example, to evaluate the plausibility of knowledge claims and decide what to believe (Sandoval et al., 2014). In this study, we concentrate on students' beliefs in justifications for knowing in the Internet context.

Bråten et al. (2005) were the first to investigate knowledge and knowing on the Internet by drawing on Hofer's and Pintrich's four dimensions of epistemic beliefs. In their study, the justification for knowing dimension ranged from the view that claims on the Internet can be accepted without critical evaluation to the view that these claims should be verified against other sources, reason, or prior knowledge. It was found that justification for knowing formed a separate dimension from the other three knowledge dimensions (See above). Thereafter, several studies have confirmed that the justification for knowing dimension is distinct from the knowledge dimensions in the Internet context (e.g., Kammerer et al., 2013; Strømsø & Bråten, 2010). Some studies have also found an association between individuals' beliefs in the justification for knowing and their critical evaluation of online information (e.g., Kammerer et al., 2013; Knight et al., 2017).

Whereas Bråten et al. (2005) examined the justification for knowing as a unidimensional construct, Greene et al. (2008) later argued that justifications for knowing cannot be captured by a single dimension. Following this assertion, Greene et al. (2008) suggested two justification for knowing dimensions: justification by authority and personal justification. Further, an additional dimension, justification by multiple sources, emerged in the think-aloud study by Ferguson et al. (2012). Kammerer et al. (2015) used a two-dimensional knowing construct including personal justification and justification by multiple sources in the Internet context. Their results showed that the more participants believed that claims need to be checked against other sources, the more time they spent on credible websites during a Web search, whereas the more they believed that claims need to be checked based on reason or prior knowledge, the more time they spent on less credible websites.

To measure the three dimensions of knowing in the Internet context, Bråten et al. (2019) developed and validated an Internet-specific Epistemic Justifications (ISEJ) inventory. It measures readers' beliefs in the evaluation of online information based on one's prior knowledge and reasoning (personal justification), on the competency and expertise of the source (justification by authority) and on checking and comparing several information sources (justification by multiple sources). A recent think-aloud study (Kammerer et al., 2021) used ISEJ among university students to examine the role of students' epistemic justifications in their source evaluation and corroboration during a Web search on a socio-scientific issue. The study showed that the more students believed that they use justification by authority the more they evaluated sources. Beliefs in personal justification were negatively associated with comments regarding corroboration of information across online texts. Further, beliefs in justification by multiple sources did not predict students' source evaluations or use of corroboration during Web search but positively predicted the quality of their justified recommendations.

1.4 | The present study

The present study examined upper secondary school students' abilities to evaluate the credibility of self-selected, authentic online texts during online inquiry. Students worked in a restricted Web environment and searched for information with Google Custom Search Engine to solve a problem concerning a health-related topic, either *Vaccination* or *Fats*. Primarily, we explored the associations between students' beliefs in justifications for knowing and their evaluation performance.

The specific research questions were:

RQ1. How well did students evaluate the credibility of self-selected online texts when provided with a range of online texts via Google Custom Search Engine?

RQ2. How were students' Internet-specific epistemic justifications associated with their evaluation performance when the usefulness of text selections, reading fluency and prior topic knowledge were controlled for?

RQ3. Did the associations between students' Internet-specific epistemic justifications and their evaluation performance differ according to the topic?

We controlled for the usefulness of students' text selections because the selections reflect their initial evaluation judgements (e.g., Hautala et al., 2018; Rieh, 2002). Further, recent research has shown that students' basic reading skills (e.g., Kanniainen et al., 2019; Potocki et al., 2020) contribute to their credibility evaluations and therefore, students' reading fluency was controlled for, too. As the topic and knowledge about it seem to play a role in the evaluation of online texts (e.g., Bråten, McCrudden, et al., 2018; Forzani, 2018) and in epistemic beliefs (e.g., Greene et al., 2008), we also controlled for students' prior topic knowledge.

2 | METHOD

2.1 | Participants

Participants consisted of 372 students (59% females, $M = 17.35$ years, $SD = 0.40$) from eight upper secondary schools in Finland. The study was embedded in the language arts course 'Texts and influence'. All students completed the tests and tasks, but only responses of those students who gave informed consent were used for the research purposes. If a student was underage, consent was also received from guardian/s.

2.2 | Online inquiry task

As a part of their language arts course, students conducted an online inquiry task in a web-based environment designed for research purposes. The task was to solve a health-related problem concerning either vaccination or saturated fats.

Following the previous research (e.g., Kammerer et al., 2015; Scharrer et al., 2019), we provided students with a task scenario that was related to a real-life problem. In the vaccination topic, students were presented with a request to help an expectant mother decide whether she should vaccinate her child or not. She reports receiving conflicting information about vaccines. In an NGO-sponsored public lecture, she had heard that babies should not be vaccinated because vaccines weaken resistance and cause autism. In turn, a health nurse in a maternity clinic had recommended that opposite. Similarly, in the fats topic, students were presented with a request to help a university student decide whether he should avoid saturated fats in his diet. He had also received conflicting information about saturated fats. At a book launch, it had been suggested that saturated fats protect against heart and vascular diseases and decrease blood cholesterol. A health nurse, in turn, had recommended avoiding saturated fats.

After reading the task scenario, the online inquiry task proceeded in four phases (Leu et al., 2019): (1) considering information need to solve the problem; (2) locating information with a search engine to select three online texts; (3) identifying main ideas of each selected

text and evaluating the credibility of the texts and (4) writing a justified recommendation. Each task phase began on a separate page. Students were able to move between the task phases by using forward and backward buttons. The data of this study originates from Task Phases 2 (selections) and 3 (credibility evaluations).

In Task Phase 2, students were asked to select three online texts with Google Custom Search Engine to provide credible information to the expectant mother or the university student. Google Custom Search Engine included 35 authentic online texts (per topic) that varied in their usefulness for the task (See Section 2.5.1 descriptions and scoring the texts). We used Google Custom Search Engine for two reasons. First, it is based on Google's core search technology and provides an authentic search experience for students. Second, it allows the inclusion of pre-selected online texts in the search engine.

Figure 1 presents the task interface for Task Phase 2. The interface was split into two areas for searching (left-hand side) and instructions and recording the response (right-hand side). By using the custom search engine, students could open as many pages as they wanted from the search results. After leaving the task phase, students were not able to change their selections.

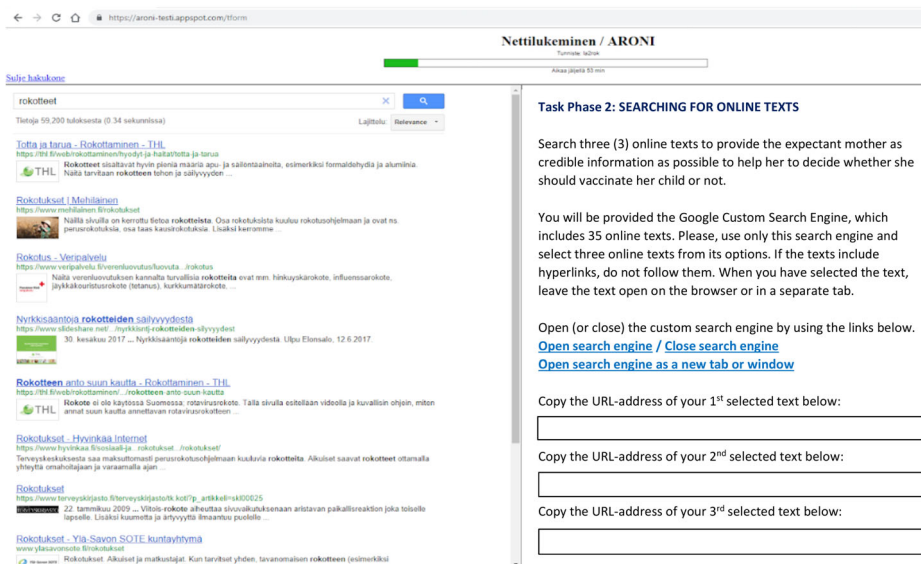
In Task Phase 3, students were asked to identify the main ideas of each text and to evaluate the credibility of the texts (See Figure 2). The URL address of the selected text was available when answering the questions, and by clicking it, students were able to open and read the whole text in a separate tab. To evaluate the credibility of the texts, students were asked to respond to two questions: *What aspects make the online text credible? What aspects may weaken the credibility of the online text?*

The latter question was supposed to facilitate students not only to confirm the credibility but also to approach the texts critically. As the online texts were authentic, they included a different amount of information about sources. For example, many texts lacked information about the author. By prompting students to also consider aspects that may weaken the credibility, we provided more equal opportunities for students to get credit from paying attention to the author, that is, either by notifying the author or by notifying the lack of author information (See Section 2.5.2 for scoring).

2.3 | Other measures

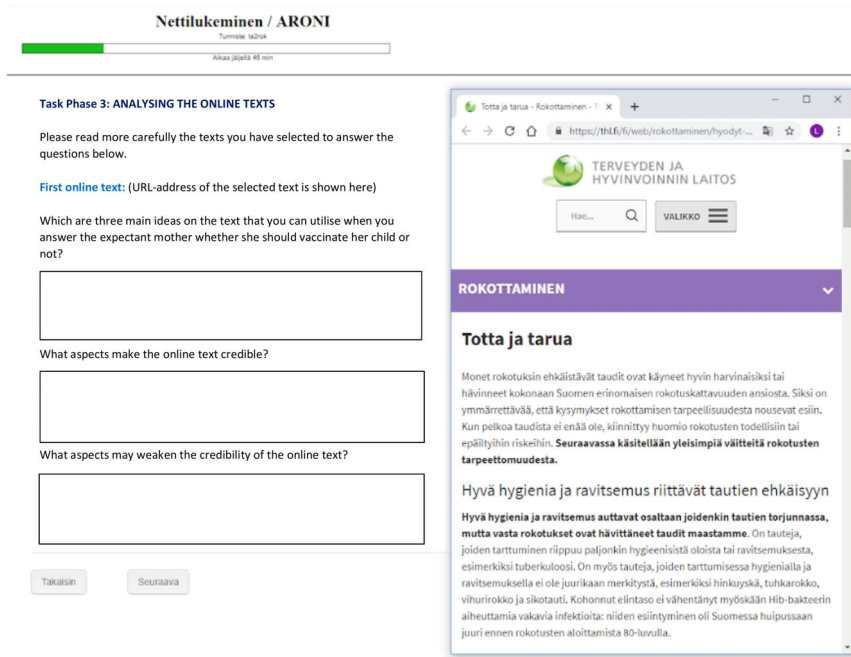
To measure students' beliefs in their justifications for knowing on the Internet context, we applied the *Internet-Specific Epistemic Justifications* (ISEJ) inventory, which has been validated with Norwegian pre-service teachers (Bråten et al., 2019). The measure was translated and adapted for Finnish upper secondary school students. When the original measure was contextualized for educational topics, our version referred to school tasks in general. The ISEJ inventory consists of 12 Likert-scale items about students' justifications for knowing when using the Internet as a knowledge resource (Bråten et al., 2019). The inventory comprises three dimensions, each of which is measured with four items: Personal Justification (e.g., 'To check whether information related to my school task I find on the Internet is reliable,

FIGURE 1 Task phase 2: Locating and selecting online texts [Colour figure can be viewed at wileyonlinelibrary.com]



Note. Instructions are translated in English. Google Custom Search Engine was shown on the left-hand side and the instructions on the right-hand side. Remaining time for the entire task was shown at the top of the interface. (Screenshot: June 11, 2019)

FIGURE 2 Task phase 3: Identifying Main ideas and evaluating the credibility of online texts [Colour figure can be viewed at wileyonlinelibrary.com]



Note. Instructions for the sub-task and related questions presented on the left-hand side are translated in English. By clicking the URL address, students were able to read the selected online text in a separate tab while writing their responses to the boxes. (Screenshot: June 11, 2019, permission for publishing the web page “Totta ja tarua” was received from Finnish Institute for Health and Welfare, THL.)

I evaluate it in relation to my knowledge of this topic’), Justification by Authority (e.g., ‘When I read information from the Internet related to my school task, I evaluate whether this information is written by an expert’) and Justification by Multiple Sources (e.g., ‘To determine whether the information related to my school task I find on the Internet is trustworthy, I compare information from multiple sources’). Instead of using the original 10-point scale, we used a 5-point scale

with labels: 1 = strongly disagree, 2 = partly disagree, 3 = not disagree or agree, 4 = partly agree, 5 = strongly agree. Thus, the ISEJ-items were measured on the ordinal level and used as approximations of students' continuous level beliefs in justifications for knowing.

Reading fluency was measured with a word-chain test, comprising 25 chains, each containing four words written without intervening spaces (Holopainen et al., 2004). Students were asked to separate as

many chains into primary words as possible within 90 s. The total score was the number of correctly separated words (0–100). According to the test manual, the test–retest reliability coefficient of the test varied between 0.70 and 0.84.

Prior topic knowledge measure comprised 10 statements, three correct and seven incorrect, on either vaccination or fats. Students were asked to select the three statements they considered correct. They earned one point for each correct statement or non-selected incorrect statement (0 or 1 per statement). Four items on each topic were excluded because they were either too easy or too difficult. Hence, the maximum score for each topic was 6 points. Reliability was 0.66 with 95% CI [0.53–0.79] for vaccination and 0.83 with 95% CI [0.66–0.99] for fats (Raykov et al., 2010).

2.4 | Procedure

Students filled in the ISEJ inventory before the research session and returned it to the teacher. The research session was conducted during a 75-min lesson in classrooms. Before the online inquiry task, students were administered a reading fluency test. They then accessed the Web-based environment with a code and performed the prior topic knowledge test and the online inquiry task. The researcher randomly allocated the code for the vaccination topic to half of the students and the code for the fats topic to the other half. Students had 60 min to complete the entire online inquiry task. The researcher gave the students instructions and helped if they encountered technical problems.

2.5 | Data analysis

2.5.1 | Selection of online texts

In Task Phase 2 (Figure 1), students selected three online texts by using Google Custom Search Engine that included 35 pre-selected texts per topic. Although instructed to select only texts included in the custom search engine, one-fourth of students also selected other texts, mostly one. These other texts ($N = 64$) accounted for 11% of all

selected texts (Table 1). Almost 60% of these texts appeared in the same venue as the pre-selected texts suggesting that students probably navigated within the website. We incorporated the other selected texts into the original textbase and used the same rubric to score all 134 texts.

In scoring, we applied the framework of the text usefulness by McCrudden (2018, p. 179) including two dimensions: text relevance (more-relevant vs. less-relevant) and source credibility (higher vs. lower source credibility). For our analysis, we added the third level to both of these dimensions: 'irrelevant' for the text relevance dimension and 'not credible' for the source credibility dimension.

By utilizing these dimensions, we established four categories of text usefulness: (1) More useful texts (more-relevant texts with higher source credibility), (2) Useful texts (more-relevant texts with lower source credibility AND less-relevant texts with higher source credibility), (3) Less useful texts (less-relevant texts with lower source credibility) and (4) Not useful texts (irrelevant AND/OR not credible texts) (See Appendix S1). The texts were classified based on the first and second authors' shared discussions about their relevance and credibility. As students were asked to select three online texts, the maximum score for their selections was nine points. Table 1 presents the number of texts that were classified into each of the categories and proportion of students' text selections.

2.5.2 | Students' credibility evaluations

In Task Phase 3, students answered the questions: *What aspects make the text credible?* and *What aspects may weaken the credibility of the text?* We considered these responses as one unit of analysis for each self-selected online text. The analysis proceeded in two steps. In Step 1, we examined how students evaluated each text in terms of different aspects of credibility. In Step 2, we utilized the results of Step 1 to assess students' evaluation performance across all three selected texts.

Step 1: Aspects of credibility. In our analysis, we focused on central aspects of the evaluation of credibility: evaluation of the source of the online texts, more precisely the author, venue and their intentions (e.g., Bråten, Stadler, et al., 2018), evaluation of evidence (Forzani, 2020; Sinatra & Lombardi, 2020) and corroboration

TABLE 1 Number of pre-selected and other selected texts by topic and proportion of all selections, presented according to texts' usefulness

Category	Number of pre-selected texts ($N = 35$ per topic)		Number of other selected texts ($N = 64$)		Total	Proportion (%) of all selections ($N = 1031$) ^a	
	Vaccination	Fats	Vaccination	Fats		Pre-selected	Other selected
More useful texts (3 points)	3	3	5	4	15	54	1
Useful texts (2 points)	5	5	8	11	29	24	3
Less useful texts (1 point)	5	5	11	2	23	7	5
Not useful texts (0 points)	22	22	16	7	67	6	2
Total	35	35	40	24	134	89	11

^aStudents ($N = 345$) selected three online texts except for one student who only selected two texts (vaccination) and one student who did not select any texts (fats).

(Kohnen & Mertens, 2019; Wineburg, 1991). As argued by Forzani (2020), triangulation across different credibility aspects assists students to gain a fuller understanding of the credibility of an online text. Abilities to evaluate various credibility aspects allow students the flexibility to apply different evaluation criteria depending on the text under exploration. In addition, abilities to engage in a deep level of reasoning are pivotal (Coiro et al., 2015; Kiili et al., 2019). Given this, we created the scoring system presented in Table 2. The responses for each self-selected online text were scored for five aspects: author, venue, intentions, evidence and corroboration. Students earned 0–3 points for each aspect depending on the depth of the evaluations in their responses.

The inter-rater reliability was examined by having the first and second authors to score 10% of responses (37 students' evaluations for three online texts, altogether 111 responses). The Kappa value was calculated for each of the scored aspects and it varied from 0.78 to 0.90. The first authors' scores were used in further analysis.

Step 2: Evaluation performance. To assess students' evaluation performance across three online texts, we created a scoring rubric that utilized the analysis conducted in Step 1. The scoring rubric, presented in Table 3, acknowledged different credibility aspects and depth in students' reasoning (justifications at the highest, 3 points level). The

scoring rubric reflected whether students' responses across the three texts demonstrated their abilities to evaluate different credibility aspects and engage in deep reasoning (See also Kiili et al., 2019). In other words, students had three possibilities to evaluate each aspect, and they were given credit in the scoring system if they evaluated the aspect at least once. By this procedure, we tried to minimize the effect of the evaluation of different text combinations.

To examine the inter-rater reliability for the evaluation performance score, we used the first and second author's scores of credibility aspects (See Step 1) to calculate the evaluation performance scores. The correlation between the evaluation performance scores was 0.95. The first authors' scores were used in further analysis.

2.5.3 | Statistical analyses

Statistical analyses for RQ2 and RQ3 were conducted by using the Mplus statistical package (version 7.4; Muthén & Muthén, 1998–2017). We estimated model parameters with the maximum likelihood estimation with non-normality robust standard errors (MLR), as the ISEJ items were skewed (Appendix S2). Because missing data (range 0%–2.4%) were completely random (Little's MCAR test result:

TABLE 2 Scoring for credibility aspects in students' evaluations of selected online texts

Aspect	0 point	1 point	2 points	3 points
Author	Student does not refer to any evaluation criteria related to the author.	Student refers to author without mentioning her/his name or any author-related source features (e.g., author is an expert) OR student notices that author is not mentioned.	Student refers to one author-related source feature (e.g., doctor) with or without naming the author.	Student names the author AND refers to at least two author-related source features (e.g., credentials, affiliation).
Venue	Student does not refer to any evaluation criteria related to the venue.	Student refers to publication practices without specifying them or naming the venue (e.g., experts write to this website).	Student names the venue OR specifies the publication practices OR refers to venue's areas of expertise.	Student names the venue AND specifies its publication practices OR areas of expertise in a detailed way.
Intentions	Student does not refer to any evaluation criteria related to intentions.	Student refers to intentions in a general manner (e.g., objective, unbiased) OR student notices commercials or their absence.	Student refers to intentions with some specification (e.g., organization has no commercial purposes).	Student describes intentions in a detailed way (e.g., organization investigates public health and makes efforts to promote it).
Evidence	Student does not refer to any evaluation criteria related to evidence.	Student refers to evidence in a general manner (e.g., references/statistics are provided OR not provided).	Student refers to evidence with some specification (e.g., includes research-based information/medical knowledge).	Student describes evidence in a detailed way (e.g., the interviewed doctor is a head of vaccination department from National Institute for Health and Welfare).
Corroboration	Student does not refer to corroboration as an evaluation criteria.	Students refers to the teacher recommendation OR previous experiences with the website OR notifies that information could be corroborated.	Student mentions that similar issues appear in other texts without specifying those sources.	Student explicitly corroborates the information by linking two or more of the selected online texts.

TABLE 3 Scoring for students' evaluation performance across three online texts and amount of students (*f*, %) in the categories

Score	Criteria	Vaccination		Fats		All	
		<i>f</i>	%	<i>f</i>	%	<i>f</i>	%
0	Students does not evaluate any of the five credibility aspects.	2	1.2	1	0.6	3	0.9
1	Student evaluates 1 aspect of the credibility but not at the highest quality level.	20	12.0	8	4.5	28	8.1
2	Student evaluates 1 aspect of the credibility with one or two evaluations at the highest quality level OR Student evaluates 2 aspects of the credibility but not at the highest quality level.	38	22.8	30	16.9	68	19.7
3	Student evaluates 2 aspects of the credibility with at least one evaluation at the highest quality level OR Student evaluates 3–4 aspects of the credibility but not at the highest quality level.	51	30.5	66	37.0	117	33.9
4	Student evaluates 3–4 aspects of the credibility with one or two evaluations at the highest quality level.	45	26.9	57	32.0	102	29.6
5	Student evaluates 3–4 aspects of the credibility with at least three evaluations at the highest quality level OR Student evaluates 5 aspects of the credibility with at least one evaluation at the highest quality level.	11	6.6	16	9.0	27	7.8
Total		167	100.0	178	100.0	345	100.0

$\chi^2[78] = 84.72, p = 0.28$; Little, 1988), we used the full information maximum likelihood procedure to account for missing data (Enders, 2010). In the data, students were nested within courses. Although intra-class correlations at the course level were small (range 0.00–0.10), we used a course as a clustering variable and estimated unbiased standard errors by using the COMPLEX option.

We examined associations between students' Internet-Specific Epistemic Justifications (ISEJ) and their evaluation performance via structural equation modelling (SEM) (Figure 3). In the model, Evaluation Performance was the dependent variable and the three justification for knowing dimensions were independent variables. Reading Fluency, Prior Topic Knowledge and Selection Score were controlled for.

Before the main analyses, we examined via CFA whether our data confirmed the original three-dimensional structure of the ISEJ inventory (See Appendix S3). As the dimensions were multicollinear (range of correlations 0.57–0.66), we used hierarchical regression analysis within the SEM framework to examine the unique effects of the knowing dimensions on Evaluation Performance. This enabled us to separate the unique variance of each dimension from the shared variance between the three dimensions via the Cholesky factoring approach (de Jong & van der Leij, 1999).

Cholesky factoring for the justification for knowing dimensions (Figure 3) was performed so that we set the first Cholesky factor (labelled 'PJ: Cholesky') to explain all the variance unique to the Personal Justification dimension and the variance it shares with the other two dimensions. Then, we set the second Cholesky factor (labelled 'JA: Cholesky') to explain the unique variance of the Justification by Authority dimension and the variance it shares with Justification by Multiple Sources. The third Cholesky factor (labelled 'JMS: Cholesky') captured the remaining (unique) variance of the JMS dimension. The correlations between the Cholesky factors and the

correlations between the original justification for knowing dimensions and their cross-correlations were fixed to 0.

We entered PJ first because it can be regarded as a more simplistic epistemic justification belief for non-experts than JA and JMS (cf. Bromme & Goldman, 2014). In addition, JA and JMS reflect the evaluation practices that are central to the documents model framework (Britt et al., 2018). JMS was entered last because it reflects the evaluation practices of experts (Kohnen & Mertens, 2019; Wineburg, 1991) that are more rarely observed among students compared to practices reflecting JA (e.g., Kammerer et al., 2021; Kiili et al., 2019).

Next, we regressed Evaluation Performance on the Cholesky factors in a hierarchical order determined by the formation process of the Cholesky factors (de Jong & van der Leij, 1999). First, we set the PJ Cholesky factor to explain Evaluation Performance. Then, we set JA Cholesky factor to explain the remaining variance of Evaluation Performance (i.e., variance not explained by the PJ Cholesky factor). Thereafter, the JMS Cholesky factor was set to explain the remaining variance of Evaluation Performance.

Finally, we examined topic differences in the linkages between the Cholesky factors and Evaluation Performance by using the multigroup procedure (Figure 3). The fit of the freely estimated model was compared to that of the constrained model by using the Satorra-Bentler χ^2 difference test (Satorra & Bentler, 2001).

We evaluated the goodness-of-fit of all the tested CFA and SEM models with the χ^2 test. However, as the χ^2 test is sensitive to the non-normality of data and model complexity, we evaluated the model fit also with the Root-Mean-Square of Approximation (RMSEA) with a 90% confidence interval, Comparative Fit Index (CFI), Tucker-Lewis Index (TLI) and Standardized Root-Mean-Square Error (SRMR). Values indicating good model fit are as follows: χ^2 test $p > 0.05$, RMSEA < 0.06 , CFI and TLI > 0.95 and SRMR < 0.08 (Hu & Bentler, 1999).

Before the analysis of topic differences in the linkages between the Cholesky factors and Evaluation Performance, we investigated the invariance of the ISEJ measurement model across topics (Meredith, 1993) (See Appendix S3) by using the Satorra-Bentler χ^2 difference test (Satorra & Bentler, 2001). A statistically non-significant χ^2 difference test denotes that the model with more invariance constraints fits the data better than the model with fewer invariance constraints. However, because the χ^2 test is sensitive to the non-normality of variables, we also used the CFI, RMSEA and SRMR criteria (Chen, 2007). A change (Δ) below -0.01 in CFI supplemented by Δ RMSEA <0.015 and Δ SRMR <0.03 (Chen, 2007) indicates that the hypothesis of invariance should not be rejected, even if the χ^2 difference test indicates otherwise.

3 | RESULTS

3.1 | Credibility evaluations

3.1.1 | Evaluation of credibility aspects

Table 4 shows that students most often evaluated the venue and evidence presented in online texts. Specifically, almost 90% of students evaluated the venue and over 75% the evidence at least once across three online texts. In contrast, students only sparsely evaluated intentions or applied corroboration as a credibility evaluation criterion.

Furthermore, the students most often reached the highest level in their evaluations when they evaluated the evidence or venue. Over one-fourth of the students evaluated the evidence at least once at the highest level across three texts and correspondingly, one-fifth of the students when evaluating the venue.

3.1.2 | Evaluation performance

On average, students scored 3.07 for their evaluation performance (Appendix S4). As Table 3 shows, over one-third (37.4%) of students demonstrated a high ability to evaluate the credibility of online texts and one-third (33.9%) of students performed at the average level. However, almost one-tenth of students performed very poorly (0.9% scored 0 points and 8.1% scored 1 point). An additional 19.7% of students also demonstrated having limited evaluation skills. Students who explored fats scored statistically significantly higher (3.22, $SD = 1.02$) than students who explored vaccination (2.90, $SD = 1.16$) (Appendix S4).

3.2 | Associations between internet-specific epistemic justifications and evaluation performance

Figure 4 presents the results for the associations between Cholesky factors for Internet-Specific Epistemic Justifications and Evaluation

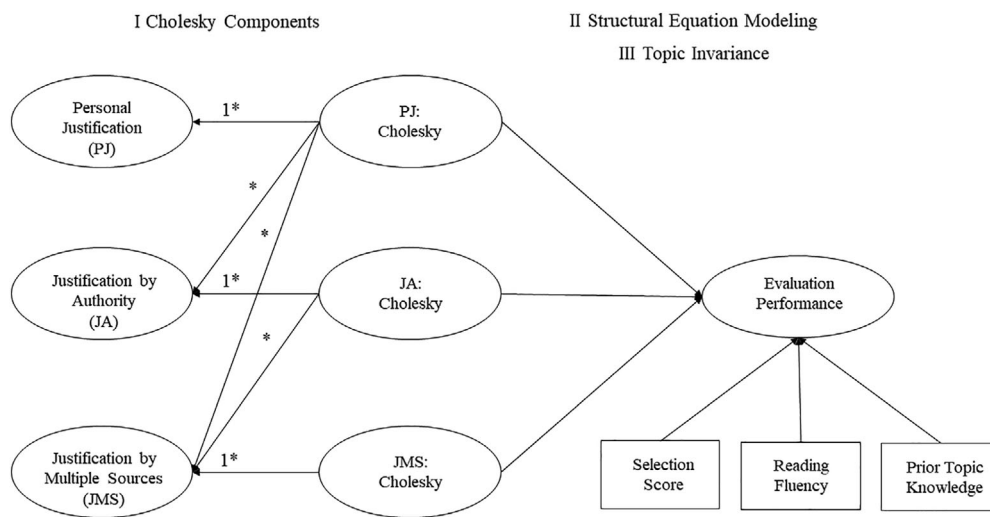


FIGURE 3 Conceptual model of the relationships between ISEJ-factors, evaluation performance and control variables. 1* fixed to one, * freely estimated

TABLE 4 Descriptive statistics for students' ($N = 345$) evaluations of credibility aspects

Aspect (range 0-9)	M (SD)	Students who evaluated the aspect at least once	
		Across three texts f (%)	At the highest-level f (%)
Venue	3.85 (2.18)	308 (89.3)	70 (20.3)
Evidence	2.72 (2.28)	265 (76.8)	91 (26.4)
Author	1.64 (1.71)	211 (61.2)	36 (10.4)
Intentions	0.69 (1.31)	99 (28.7)	23 (6.7)
Corroboration	0.32 (0.95)	48 (13.9)	9 (2.6)

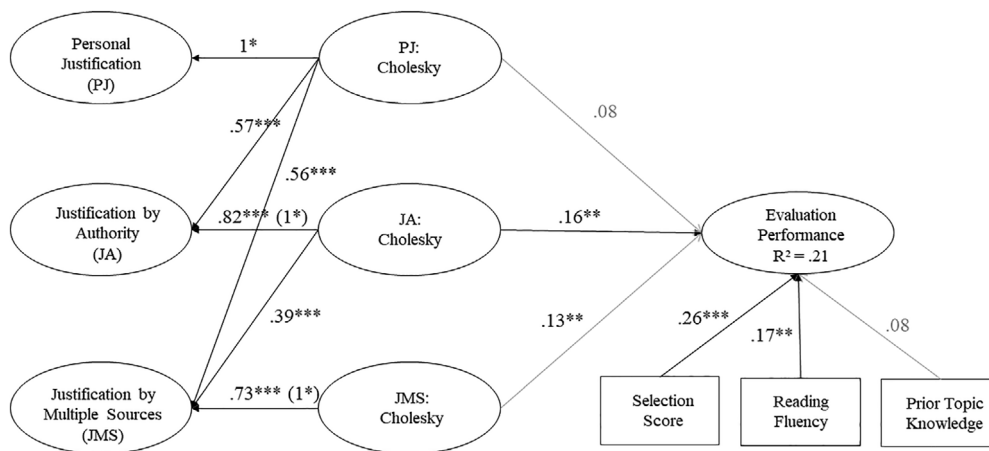


FIGURE 4 Relations between ISEJ-factors, evaluation performance and control variables. Statistically significant standardized estimates (** $p < 0.01$, *** $p < 0.001$) are written in black (nonsignificant estimates written in grey). 1* fixed to one

Performance. This model showed a good fit to the data: $\chi^2(95) = 137.24$, $p < 0.01$, CFI = 0.98, TLI = 0.97, RMSEA = 0.04 with 90% CI [0.02–0.05] and SRMR = 0.05. Of the Cholesky factors, Justification by Authority and Justification by Multiple Sources were positively associated with Evaluation Performance. Thus, students' evaluation performance was better the more they believed that they evaluate authority or/and compare multiple sources when they read online texts. Personal Justification was not associated with students' evaluation performance. The associations of Cholesky factors with Evaluation Performance were similar across the topics (RQ3): $\Delta\chi^2(6) = 5.34$, $p = 0.50$. Further, students who selected more useful texts and/or possessed better reading fluency were also better evaluators, and vice versa. Prior Topic Knowledge was only approaching statistical significance in relation to Evaluation Performance ($p = 0.10$).

4 | DISCUSSION

This study examined upper secondary school students' abilities to evaluate health-related online texts and the associations between students' beliefs in justifications for knowing and evaluation performance. The novelty of this study lies in the use of an authentic but restricted Web environment. To a limited extent, this allowed us control over the online texts that students selected and evaluated while simultaneously offering students an authentic information search experience. Our study is one of the few large-scale evaluation studies to make use of authentic online texts (See also Knight et al., 2017). We are also among the first to examine the relations of the three-dimensional Internet-Specific Epistemic Justifications in relation to students' evaluation performance (See also Kammerer et al., 2021).

The results showed considerable variation in students' abilities to evaluate the credibility of online texts, a finding also previously reported (e.g., Kiili et al., 2019; McGrew et al., 2018). Over one-third of the students demonstrated a high ability to move across different credibility aspects with some deep-level justifications when evaluating the credibility across three online texts. The ability to pay attention to

different aspects of credibility provides students with flexibility in their evaluations. Further, paying attention to multiple aspects of credibility is important, as an accurate evaluation often cannot be made by relying on one aspect alone (Forzani, 2020). Almost one-tenth of students performed very poorly, relying at most on only one aspect of credibility. Additionally, 20% of students demonstrated limited abilities to engage in versatile and sophisticated evaluation. This is worrying, as adolescents with poor evaluation skills may be particularly vulnerable to mis- and disinformation.

Of the credibility aspects, students most often evaluated venue and evidence, the latter of which has been found to be difficult for younger students (e.g., Hämäläinen et al., 2020). Over 60% of students considered the author or absence of the author information. However, students quite rarely evaluated intentions. It might be that they considered intentions of the particular authors (e.g., scientist) or publishers (e.g., an online library for medicine) to be obvious and hence did not include it in their responses. On the other hand, most of the students noticed commercials when these were included in the online texts, which is in contrast with the study by McGrew et al. (2018).

Further, students seldom used corroboration as an evaluation criterion. The infrequent use of corroboration was expected, as it is a typical expert reader strategy (Kohnen & Mertens, 2019; Wineburg, 1991). Selected combinations of texts were not, however, ideal for corroborative purposes owing to the few discrepancies between them, as discrepancies have been found to promote comparison of the content and source features of documents (e.g., Kammerer et al., 2016; Rouet et al., 2016). In addition, even though students may have purposefully selected the texts that supported each other, they did not explicate this in their responses.

We also found that the evaluation performance of students, who believed that the credibility of the information they find on the Internet needs to be justified by the expertise of the source, was higher in quality. This is in line with findings by Kammerer et al. (2021) regarding the value of students' beliefs in justification by the authority to students' evaluations of online texts. Along with this result, students' attention to author and venue is encouraging as author expertise has

been considered one of the most important source features requiring evaluation (e.g., Britt et al., 2014; Potocki et al., 2020), particularly in situations where the reader does not have much prior knowledge (Bråten, McCrudden, et al., 2018).

Furthermore, when students' epistemic justifications reflected a need for corroboration when evaluating online texts, they evaluated more carefully the credibility of selected texts. The association between students' evaluation performance and their beliefs in justification by multiple sources is in line with the Web search study of health information (Kammerer et al., 2015) but contrary to the recent ISEJ-study by Kammerer et al. (2021) regarding students' spontaneous evaluations during Web search. Our study suggests that when students are prompted to pay attention to the credibility of online texts, their beliefs in justification by multiple sources play a role in credibility evaluations. However, as our findings showed, students rarely referred to corroboration in their credibility evaluations. Thus, it seems that being aware of the importance of corroboration does not necessarily lead to its deployment in evaluation situations.

Further, students' beliefs in personal justification were not associated with their evaluation performance, not even negatively, as has been found in the studies by Kammerer et al. (2015, 2021). It should be noticed that in our study personal justification items were not context-based like in the study by Kammerer et al. (2021) which might have affected this result. That is, own prior knowledge and reasoning can be restricted especially in regard to unsettled natural science topics. Notably, comparing the information with personal knowledge is quite often an uncertain evaluation strategy, as personal knowledge can include false beliefs or biased information (Greene et al., 2019).

Finally, we also examined whether the associations between students' beliefs in justifications for knowing and their evaluation performance differed according to the topic. Interestingly, all three associations were similar in both topics (vaccination and fats), although the students whose topic was fats performed better in the prior topic knowledge test and in selecting and evaluating of online texts than those whose topic was vaccination. These results suggest that the newly developed measure for Internet-Specific Epistemic Justifications validated with pre-service teachers (Bråten et al., 2019) is also valid for use among upper secondary school students and with different health topics.

4.1 | Limitations and future research

This study has its limitations. First, despite our ambitious effort to create an authentic but restricted Web environment through Google Custom Search Engine for examining students' credibility evaluations, students also selected online texts that were not included in it. These other texts, however, accounted only for 11% of all text selections. Even though students were exposed to different text materials, the developed scoring system for credibility evaluations allowed the flexibility to assess students' evaluations across different texts.

Second, students completed the online inquiry task by following the predetermined task order. This did not allow them to engage in

iterative processes typical for online inquiry (e.g., Rieh, 2002). For example, when evaluating the self-selected texts students were not able to change their selections even though they might have realized that the selected texts were not the best possible to solve the problem. However, examining online inquiry as an iterative process adds complexities that are quite difficult to handle with a large sample size ($N = 372$) that we had in this study.

Third, in the online inquiry task, students were prompted to evaluate the credibility of online texts with specific questions facilitating their evaluations of online information that may otherwise be rare (Gerjets et al., 2011; Paul et al., 2017). Thus, our results reflect what students are capable of doing, and not necessarily, how they spontaneously engage in the evaluation of online information. We decided to use prompts because the understanding of students' strategic repertoire provides valuable information for developing instruction.

Fourth, because we scored students' evaluation performance holistically, covering both the evaluation of different credibility aspects and depth of reasoning, we were unable to measure the association between single credibility aspects (e.g., corroboration) and particular justifications for knowing (e.g., justification by multiple sources). Examination of the associations of the different credibility aspects with students' justifications for knowing would have better revealed how realistically students believed that they were evaluating online information by using specific evaluation criteria. Based on previous research (e.g., Paul et al., 2017), it is known that students tend to overestimate their skills; in the present study, their self-evaluations reflected rather positive beliefs about their evaluation behaviour. These specific associations could be investigated in future studies.

4.2 | Instructional implications

The present results indicate a need for instruction that addresses both, evaluation of different credibility aspects and depth in evaluations. Instruction that combines the different credibility aspects emphasized in this study could enhance evaluation. It is important to discuss with students why multiple aspects should be evaluated and to point out that an evaluation based on one aspect alone could be misleading. For example, claims made in a blog post written by a layperson and an expert may vary in plausibility. In addition, personal feedback could help students to view their abilities more realistically and promote advanced justifications for knowing that, in turn, can positively influence their intertext model construction (Bråten et al., 2011). The value of corroboration as an expert strategy (e.g., Kohnen & Mertens, 2019) could also be highlighted in instruction. While students believed that they often corroborate online information, this was not confirmed by their evaluation performance. Corroboration is of particular importance in building a coherent understanding of the topic in question (cf. Perfetti et al., 1999).

Given that some students are already skilled evaluators, teachers could apply collaborative learning methods whereby students can share effective evaluation strategies and learn from each other (e.g., Kiili et al., 2019). Such collaborative learning could be organized

around a scripted online inquiry process in different disciplines. As evaluation occurs during different phases of online inquiry (e.g., Leu et al., 2019), it could be practised during several consecutive lessons focusing on one process at a time. To design successful collaborative learning experiences for students, collaboration needs to be supported (e.g., Jeong & Hmelo-Silver, 2016). One way to support collaboration is to use shared working templates, which include prompts that support students to critically search, select, evaluate and synthesize online information. A recent review (Cartiff et al., 2020) reported that guided forms of instruction and models emphasizing justification and source evaluation are effective in promoting students' epistemic cognition and academic achievement.

ACKNOWLEDGMENTS

The authors would like to thank Timo Salminen, Minna Lakkala, Liisa Iilomäki and Tuulikki Alamettälä for the data collection, Jari Hämäläinen for the technology design, Michael Freeman for his valuable comments on the language, and the students who participated in the study. This research was funded by the Academy of Finland (Project Number: 285806).

CONFLICT OF INTEREST

The authors declare no potential conflicts of interest with respect to the research, authorship and/or publication of this article.

PEER REVIEW

The peer review history for this article is available at <https://publons.com/publon/10.1111/jcal.12580>.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are available from the corresponding author upon reasonable request.

ORCID

Elina K. Hämäläinen  <https://orcid.org/0000-0001-7561-0530>

Carita Kiili  <https://orcid.org/0000-0001-9189-4094>

Eija Räikkönen  <https://orcid.org/0000-0003-4450-9178>

Miika Marttunen  <https://orcid.org/0000-0002-4554-9764>

REFERENCES

- Braasch, J. L. G., Bråten, I., Strømsø, H. I., Anmarkrud, Ø., & Ferguson, L. E. (2013). Promoting secondary school students' evaluation of source features of multiple documents. *Contemporary Educational Psychology*, 38, 180–195. <https://doi.org/10.1016/j.cedpsych.2013.03.003>
- Brante, E. W., & Strømsø, H. I. (2018). Sourcing in text comprehension: A review of interventions targeting sourcing skills. *Educational Psychology Review*, 30, 773–799. <https://doi.org/10.1007/s10648-017-9421-7>
- Bråten, I., Brandmo, C., & Kammerer, Y. (2019). A validation study of the Internet-Specific Epistemic Justification Inventory with Norwegian preservice teachers. *Journal of Educational Computing Research*, 57, 877–900. <https://doi.org/10.1177/0735633118769438>
- Bråten, I., Britt, M. A., Strømsø, H. I., & Rouet, J.-F. (2011). The role of epistemic beliefs in the comprehension of multiple expository texts: Toward an integrated model. *Educational Psychologist*, 46, 48–70. <https://doi.org/10.1080/00461520.2011.538647>
- Bråten, I., McCrudden, M. T., Stang Lund, E., Brante, E. W., & Strømsø, H. I. (2018). Task-oriented learning with multiple documents. Effects of topic familiarity, author expertise, and content relevance on document selection, processing, and use. *Reading Research Quarterly*, 53, 345–365. <https://doi.org/10.1002/rrq.197>
- 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.), *Handbook of discourse processes* (2nd ed.). Routledge.
- Bråten, I., Strømsø, H. I., & Samuelstuen, M. S. (2005). The relationship between internet-specific epistemological beliefs and learning within internet technologies. *Journal of Educational Computing Research*, 33, 141–171. <https://doi.org/10.2190/E763-XOLN-6NMF-CB86>
- 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*, 4, 104–122. <https://doi.org/10.1080/00461520.2014.916217>
- Britt, M. A., Rouet, J. F., & Durik, A. (2018). Representations and processes in multiple source use. In J. L. G. Braasch, I. Bråten, & M. T. McCrudden (Eds.), *Handbook of multiple source use* (pp. 17–33). Routledge.
- Bromme, R., & Goldman, S. R. (2014). The public's bounded understanding of science. *Educational Psychologist*, 49, 59–69. <https://doi.org/10.1080/00461520.2014.921572>
- Cartiff, B. M., Duke, R. F., & Greene, J. A. (2020). The effect of epistemic cognition interventions on academic achievement: A meta-analysis. *Journal of Educational Psychology*, 113, 477–498. <https://doi.org/10.1037/edu0000490>
- Chen, F. F. (2007). Sensitivity of goodness of fit indexes to lack of measurement invariance. *Structural Equation Modeling*, 14, 464–504. <https://doi.org/10.1080/10705510701301834>
- Coiro, J., Coscarelli, C., Maykel, C., & Forzani, E. (2015). Investigating criteria that seventh graders use to evaluate the quality of online information. *Journal of Adolescent & Adult Literacy*, 59, 287–297. <https://doi.org/10.1002/jaal.448>
- de Jong, P. F., & van der Leij, A. (1999). Specific contributions of phonological abilities to early reading acquisition: Results from a Dutch latent variable longitudinal study. *Journal of Educational Psychology*, 91, 450–476. <https://doi.org/10.1037/0022-0663.91.3.450>
- Enders, C. K. (2010). *Applied missing data analysis*. Guilford Publications.
- Ferguson, L. E., Bråten, I., & Strømsø, H. I. (2012). Epistemic cognition when students read multiple documents containing conflicting scientific evidence: A think-aloud study. *Learning and Instruction*, 22, 103–120. <https://doi.org/10.1016/j.learninstruc.2011.08.002>
- Forzani, E. (2018). How well can students evaluate online science information? Contributions of prior knowledge, gender, socioeconomic status, and offline reading ability. *Reading Research Quarterly*, 53, 385–390. <https://doi.org/10.1002/rrq.218>
- Forzani, E. (2020). A three-tiered framework for proactive critical evaluation during online inquiry. *Journal of Adolescent & Adult Literacy*, 63, 401–414. <https://doi.org/10.1002/jaal.1004>
- Freeman, J. L., Caldwell, P. H., & Scott, K. M. (2020). The role of trust when adolescents search for and appraise online health information. *The Journal of Pediatrics*, 221, 215–223. <https://doi.org/10.1016/j.jpeds.2020.02.074>
- Gerjets, P., Kammerer, Y., & Werner, B. (2011). Measuring spontaneous and instructed evaluation processes during web search: Integrating concurrent thinking-aloud protocols and eye-tracking data. *Learning and Instruction*, 21, 220–231. <https://doi.org/10.1016/j.learninstruc.2010.02.005>
- Greene, J. A., Azevedo, R., & Torney-Purta, J. (2008). Modeling epistemic and ontological cognition: Philosophical perspectives and

- methodological directions. *Educational Psychologist*, 43, 142–160. <https://doi.org/10.1080/00461520802178458>
- Greene, J. A., Cartiff, B. M., Duke, R. F., & Deekens, V. M. (2019). A nation of curators. Educating students to be critical consumers and users of online information. In P. Kendeou, D. H. Robinson, & M. T. McCrudden (Eds.), *Misinformation and fake news in education* (pp. 187–206). Information Age Publishing.
- Greene, J. A., & Yu, S. B. (2015). Educating critical thinkers: The role of epistemic cognition. *Policy Insights From the Behavioral and Brain Sciences*, 3, 45–53. <https://doi.org/10.1177/2372732215622223>
- Hämäläinen, E. K., Kiili, C., Marttunen, M., Rääkkönen, E., González-Ibáñez, R., & Leppänen, P. H. (2020). Promoting sixth graders' credibility evaluation of web pages: An intervention study. *Computers in Human Behaviour*, 110, 106372. <https://doi.org/10.1016/j.chb.2020.106372>
- Hautala, J., Kiili, C., Kammerer, Y., Loberg, O., Hokkanen, S., & Leppänen, P. H. (2018). Sixth graders' evaluation strategies when reading internet search results: An eye-tracking study. *Behaviour & Information Technology*, 37, 761–773. <https://doi.org/10.1080/0144929X.2018.1477992>
- Hoeken, H. (2001). Anecdotal, statistical, and causal evidence: Their perceived and actual persuasiveness. *Argumentation*, 15, 425–437. <https://doi.org/10.1023/A:1012075630523>
- Hofer, B. K., & Pintrich, P. R. (1997). The development of epistemological theories: Beliefs about knowledge and knowing and their relation to learning. *Review of Educational Research*, 67, 88–140. <https://doi.org/10.3102/00346543067001088>
- Holopainen, L., Kairaluoma, L., Nevala, J., Ahonen, T., & Aro, M. (2004). *Lukivaikeuksien seulontatesti nuorille ja aikuisille [Dyslexia screening test for youth and adults]*. Niilo Mäki Institute.
- Hu, L. T., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling*, 6, 1–55. <https://doi.org/10.1080/10705519909540118>
- Jeong, H., & Hmelo-Silver, C. E. (2016). Seven affordances of computer-supported collaborative learning: How to support collaborative learning? How can technologies help? *Educational Psychologist*, 51, 247–265. <https://doi.org/10.1080/00461520.2016.1158654>
- 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. <https://doi.org/10.1016/j.chb.2015.01.045>
- Kammerer, Y., Bråten, I., Gerjets, P., & Strømsø, H. I. (2013). The role of internet-specific epistemic beliefs in laypersons' source evaluations and decisions during web search on a medical issue. *Computers in Human Behavior*, 29, 1193–1203. <https://doi.org/10.1016/j.chb.2012.10.012>
- Kammerer, Y., Gottschling, S., & Bråten, I. (2021). The role of internet-specific justification beliefs in source evaluation and corroboration during web search on an unsettled socio-scientific issue. *Journal of Educational Computing Research*, 59, 42–378. <https://doi.org/10.1177/0735633120952731>
- Kammerer, Y., Kalbfell, E., & Gerjets, P. (2016). Is this information source commercially biased? How contradictions between web pages stimulate the consideration of source information. *Discourse Processes*, 53, 430–456. <https://doi.org/10.1080/0163853X.2016.1169968>
- Kanniainen, L., Kiili, C., Tolvanen, A., Aro, M., & Leppänen, P. H. T. (2019). Literacy skills and online research and comprehension: Struggling readers face difficulties online. *Reading and Writing*, 32, 2201–2222. <https://doi.org/10.1007/s11145-019-09944-9>
- Kiili, C., Coiro, J., & Rääkkönen, E. (2019). Students' evaluation of information during online inquiry: Working individually or in pairs. *Australian Journal of Language and Literacy*, 42, 167–183.
- Kiili, C., Leu, D. J., Marttunen, M., Hautala, J., & Leppänen, P. H. T. (2018). Exploring early adolescents evaluation of academic and commercial online resources related to health. *Reading and Writing*, 31, 533–557. <https://doi.org/10.1007/s11145-017-9797-2>
- Knight, S., Rienties, B., Littleton, K., Mitsui, M., & Tempelaar, D. (2017). The relationship of (perceived) epistemic cognition to interaction with resources on the internet. *Computers in Human Behavior*, 73, 507–518. <https://doi.org/10.1016/j.chb.2017.04.014>
- Kohnen, A. M., & Mertens, G. E. (2019). “I'm always kind of double-checking”: Exploring the information-seeking identities of expert generalists. *Reading Research Quarterly*, 54, 279–297. <https://doi.org/10.1002/rrq.245>
- Leu, D. J., Kinzer, C. K., Coiro, J., Castek, J., & Henry, L. A. (2019). New literacies: A dual level theory of the changing nature of literacy, instruction, and assessment. In D. E. Alvermann, N. J. Unrau, M. Sailors, & R. B. Ruddell (Eds.), *Theoretical models and processes of literacy* (7th ed., pp. 319–346). Taylor & Francis.
- Little, R. J. (1988). A test of missing completely at random for multivariate data with missing values. *Journal of the American Statistical Association*, 83, 1198–1202. <https://doi.org/10.1080/01621459.1988.10478722>
- McCrudden, M. T. (2018). Text relevance and multiple-source use. In J. L. G. Braasch, I. Bråten, & M. T. McCrudden (Eds.), *Handbook of multiple source use* (pp. 168–183). Routledge.
- McGrew, S., Breakstone, J., Ortega, T., Smith, M., & Wineburg, S. (2018). Can students evaluate online sources? Learning from assessments of civic online reasoning. *Theory & Research in Social Education*, 46, 165–193. <https://doi.org/10.1080/00933104.2017.1416320>
- Meola, M. (2004). Chucking the checklist: A contextual approach to teaching undergraduates web-site evaluation. *Libraries and the Academy*, 4, 331–344. <https://doi.org/10.1353/pla.2004.0055>
- Meredith, W. (1993). Measurement invariance, factor analysis and factorial invariance. *Psychometrika*, 58, 525–543. <https://doi.org/10.1007/BF02294825>
- Muthén, L., & Muthén, B. O. (1998–2017). *Mplus user's guide* (8th ed.). Muthén & Muthén.
- Pan, B., Hembrooke, H., Joachims, T., Lorigo, L., Gay, G., & Granka, L. (2007). In Google we trust: Users' decisions on rank, position, and relevance. *Journal of Computer-Mediated Communication*, 12, 801–823. <https://doi.org/10.1111/j.1083-6101.2007.00351.x>
- Paul, J., Macedo-Rouet, M., Rouet, J. F., & Stadler, M. (2017). Why attend to source information when reading online? The perspective of ninth grade students from two different countries. *Computers & Education*, 113, 339–354. <https://doi.org/10.1016/j.compedu.2017.05.020>
- Perfetti, C. A., Rouet, J.-F., & Britt, M. A. (1999). Towards a theory of documents representation. In H. van Oostendorp & S. Goldman (Eds.), *The construction of mental representations during reading* (pp. 99–122). Erlbaum.
- Potocki, A., de Pereyra, G., Ros, C., Macedo-Rouet, M., Stadler, M., Salmerón, L., & Rouet, J.-F. (2020). The development of source evaluation skills during adolescence: Exploring different levels of source processing and their relationships. *Journal for the Study of Education and Development*, 43, 19–59. <https://doi.org/10.1080/02103702.2019.1690848>
- Raykov, T., Dimitrov, D. M., & Asparouhov, T. (2010). Evaluation of scale reliability with binary measures using latent variable modeling. *Structural Equation Modeling*, 17, 265–279. <https://doi.org/10.1080/10705511003659417>
- Rieh, S. Y. (2002). Judgment of information quality and cognitive authority in the web. *Journal of the American Society for Information Science and Technology*, 53, 145–161. <https://doi.org/10.1002/asi.10017>
- Rouet, J. F. (2006). *The skills of document use: From text comprehension to web-based learning*. Erlbaum.
- Rouet, J. F., Le Bigot, L., de Pereyra, G., & Britt, M. A. (2016). Whose story is this? Discrepancy triggers readers' attention to source information in short narratives. *Reading and Writing*, 29, 1549–1570. <https://doi.org/10.1007/s11145-016-9625-0>
- Salmerón, L., Strømsø, H. I., Kammerer, Y., Stadler, M., & van den Broek, P. (2018). Comprehension processes in digital reading. In M. Barzillai, J. Thomson, S. Schroeder, & P. van den Broek (Eds.), *Learning*

- to read in a digital world (pp. 91–120). Studies in written language and literacy 17). John Benjamins Publishing Company.
- Sandoval, W. A., Sodian, B., Koerber, S., & Wong, J. (2014). Developing children's early competencies to engage with science. *Educational Psychologist*, 49, 39–152. <https://doi.org/10.1080/00461520.2014.917589>
- Satorra, A., & Bentler, P. M. (2001). A scaled difference chi-square test statistic for moment structure analysis. *Psychometrika*, 66, 507–514. <https://doi.org/10.1007/BF02296192>
- Scharrer, L., Stadtler, M., & Bromme, R. (2019). Judging scientific information: Does source evaluation prevent the seductive effect of text easiness? *Learning and Instruction*, 63, 101215. <https://doi.org/10.1016/j.learninstruc.2019.101215>
- Sinatra, G. M., & Lombardi, D. (2020). Evaluating sources of scientific evidence and claims in the post-truth era may require reappraising plausibility judgments. *Educational Psychologist*, 55, 120–131. <https://doi.org/10.1080/00461520.2020.1730181>
- Strømsø, H. I., & Bråten, I. (2010). The role of personal epistemology in the self-regulation of internet-based learning. *Metacognition Learning*, 5, 91–111. <https://doi.org/10.1007/s11409-009-9043-7>
- Wineburg, S. (1991). Historical problem solving: A study of the cognitive processes used in the evaluation of documentary and pictorial evidence. *Journal of Educational Psychology*, 83, 73–87. <https://doi.org/10.1037/0022-0663.83.1.73>

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of this article.

How to cite this article: Hämäläinen, E. K., Kiili, C., Rääkkönen, E., & Marttunen, M. (2021). Students' abilities to evaluate the credibility of online texts: The role of internet-specific epistemic justifications. *Journal of Computer Assisted Learning*, 1–14. <https://doi.org/10.1111/jcal.12580>