

**This is a self-archived version of an original article. This version may differ from the original in pagination and typographic details.**

**Author(s):** Hämäläinen, Elina K.; Kiili, Carita; Räikkönen, Eija; Lakkala, Minna; Ilomäki, Liisa; Toom, Auli; Marttunen, Miika

**Title:** Teaching sourcing during online inquiry – adolescents with the weakest skills benefited the most

**Year:** 2023

**Version:** Published version

**Copyright:** © The Author(s) 2022

**Rights:** CC BY 4.0

**Rights url:** <https://creativecommons.org/licenses/by/4.0/>

**Please cite the original version:**

Hämäläinen, E. K., Kiili, C., Räikkönen, E., Lakkala, M., Ilomäki, L., Toom, A., & Marttunen, M. (2023). Teaching sourcing during online inquiry – adolescents with the weakest skills benefited the most. *Instructional Science*, 51(1), 135-163. <https://doi.org/10.1007/s11251-022-09597-2>



# Teaching sourcing during online inquiry – adolescents with the weakest skills benefited the most

Elina K. Hämäläinen<sup>1</sup> · Carita Kiili<sup>2</sup> · Eija Räikkönen<sup>3</sup> · Minna Lakkala<sup>4</sup> ·  
Liisa Ilomäki<sup>4</sup> · Auli Toom<sup>5</sup> · Miika Marttunen<sup>1</sup>

Received: 14 December 2021 / Accepted: 20 July 2022  
© The Author(s) 2022

## Abstract

Sourcing - identifying, evaluating, and using information about the sources of information - assists readers in determining what to trust when seeking information on the Internet. To survive in the post-truth era, students should be equipped with sufficient sourcing skills. This study investigated the efficacy of a teacher-led intervention aimed at fostering upper secondary school students' ( $N=365$ ) sourcing during online inquiry. The intervention ( $4 \times 75$  min) was structured in accordance with the phases of online inquiry: locating, evaluating, synthesizing, and communicating information. During the intervention, teachers demonstrated why and how to source, and students practiced sourcing by investigating a controversial topic on the Internet. Students worked in small groups and their work was supported with analysis and reflection prompts. Students' sourcing skills were measured with a web-based online inquiry task before and after the intervention. Compared to controls, the intervention fostered students' abilities in three of the four skills measured (sourcing in search queries, credibility judgments, and written product). Depending on the sourcing skill, 4–25% of students showed improved performance. The students with low sourcing skills to begin with, benefited the most from the intervention. The study demonstrated that students' sourcing skills can be supported throughout online inquiry.

**Keywords** Intervention · Online inquiry · Sourcing · Adolescents · Multiple document comprehension

---

✉ Elina K. Hämäläinen  
elina.k.hamalainen@jyu.fi

<sup>1</sup> Department of Education, University of Jyväskylä, Jyväskylä, Finland

<sup>2</sup> Faculty of Education and Culture, Tampere University, Tampere, Finland

<sup>3</sup> Faculty of Education and Psychology, University of Jyväskylä, Jyväskylä, Finland

<sup>4</sup> Faculty of Educational Sciences, University of Helsinki, Helsinki, Finland

<sup>5</sup> Centre for University Teaching and Learning, Faculty of Educational Sciences, University of Helsinki, Helsinki, Finland

## Introduction

One of the more recent developmental waves in literacy education is the digital wave in which the reader is seen as an information explorer (Tierney & Pearson, 2021) who engages in online inquiry to solve problems and make meaning of various topics (Coiro, 2021; Leu et al., 2019). Online inquiry includes the processes of specifying information need and locating, critically evaluating, synthesizing, and communicating online information (Leu et al., 2019). When engaging in successful online inquiry, a skillful digital reader attends to, represents, and evaluates the sources of the information found (Bråten et al., 2018c). These practices, termed sourcing (Bråten et al., 2018c; Wineburg, 1991), assist readers to avoid trusting misleading information, which is widespread on the Internet. A recent study (Kiili et al., 2021) showed that sourcing can be employed throughout online inquiry, and readers may engage in sourcing also in the earliest phases of inquiry. Interestingly, sourcing in the earlier phases of online inquiry supported sourcing in the later phases of inquiry, suggesting the importance of approaching sourcing as an iterative practice.

Despite the importance of sourcing, studies in offline and online contexts have shown that many students lack adequate sourcing skills (e.g., Barzilai et al., 2015; Kobayashi, 2014; McGrew et al., 2018; Strømsø & Bråten, 2014). As a result, various intervention studies have been conducted on how students' sourcing might best be supported (see reviews by Brand-Gruwel & van Strien 2018; Brante & Strømsø, 2018; Bråten et al., 2018c). Teaching these skills is essential to equip students with strategies for managing diverse information in the 21st century. However, in the interventions implemented in the Internet context, sourcing skills have not been systematically taught and measured during all the phases of online inquiry. This study extends previous work by examining whether upper secondary school students' sourcing can be enhanced throughout online inquiry by a teacher-led intervention in an authentic Internet context.

## Sourcing during online inquiry

The present study on sourcing during online inquiry has been informed by two theoretical models: the Online Research and Comprehension Model (Leu et al., 2019, see also Kiili et al., 2018) and the Documents Model (Perfetti et al., 1999; Rouet, 2006). According to the Online Research and Comprehension Model (Leu et al., 2019), a problem-based online inquiry comprises five key processes: specifying information need and locating, critically evaluating, synthesizing, and communicating online information. In the model, these processes are considered to be recursive and reciprocal so that evaluation, for example, is intertwined with the other processes. The Documents Model (Perfetti et al., 1999; Rouet, 2006), initially developed in the context of interpreting historical documents, accentuates the importance of source information in building a coherent representation across multiple documents, including conflicting information. This requires readers to connect information about sources, such as authors and their expertise and intentions, to the documents' contents to compare, contrast, and evaluate multiple documents (Perfetti et al., 1999; Rouet, 2006). Ideally, sourcing occurs during all online inquiry phases (Kiili et al., 2021) when readers gradually build a coherent representation of the topic they examine. Next, we will describe how sourcing can be applied during each online inquiry phase.

The online inquiry begins with specifying the information need i.e., what kind of information is needed to solve a problem at hand. The skillful readers can make use of source information already in this phase of online inquiry. For example, they set goals that emphasize the importance of credible information on the topic of interest, and they can consider which sources provide the most reliable information (Kiili et al., 2021). These considerations can be employed when locating information with search engines (Leu et al., 2019).

When formulating search queries, skillful readers, who frame their search terms by citing reliable persons, organizations, or research-based information, can be considered to be practicing sourcing (Kiili et al., 2021). Furthermore, when skimming the search engine results page to make text selections online readers can attend to source features (e.g., in titles, URLs, or example texts) to initially evaluate the credibility and relevance of online texts (Hahnel et al., 2020; Rieh, 2002). Even though sourcing during selecting potential online texts from search engine result page has been previously examined (e.g., Gerjets et al., 2011; Haas & Unkel, 2017; Hautala et al., 2018) sourcing practices during specifying the information need and formulation of search queries have rarely been investigated (Kiili et al., 2021).

In recent years, students' sourcing has been increasingly examined in the later phases of online inquiry in relation to evaluating the credibility of online texts and using source information to synthesize and communicate information in written products (e.g., List et al., 2017; Salmerón et al., 2018; Strømsø et al., 2013). When skillful readers explore the selected online texts, they can evaluate texts' source information, including the author's expertise and intentions as well as the venue's area of expertise and publishing practices (cf. Perfetti et al., 1999; Rouet, 2006). Evaluation of sources informs the readers' judgments of the accuracy of information. The relation between the source and content evaluation is reciprocal, thus, the judgments of the content validity can also inform the judgments of source trustworthiness (Barzilai et al., 2020). However, the importance of source evaluation is highlighted if readers lack prior knowledge about the topic (Bråten et al., 2018b; Bromme & Goldman, 2014).

The last phases of online inquiry concern synthesizing and communicating information during which students complete and communicate their representation of the examined topic. The Documents Model (Perfetti et al., 1999; Britt et al., 2018) is particularly useful to understand how readers synthesize selected information in their written products. According to the Documents Model, readers can construct two types of representations when reading multiple texts: an intertext model and an integrated mental model. The intertext model posits that source information (e.g., author/venue and their expertise/intentions) is connected to the document's content and other information sources (Perfetti et al., 1999; Rouet, 2006). These links are of two different types: source-to-content and source-to-source.

Source-to-content links show how a reader combines information about the source of a document with its content whereas source-to-source links show how a reader connects sources from multiple documents by showing the relationships between them, such as supporting, complementing, or opposing. The intertext model is particularly useful in situations where readers confront conflicting information that prevents them from coherently integrating the content of multiple documents, the reliability of which needs to be ensured (Britt et al., 2014). The integrated mental model, in turn, focuses on the content of documents and describes readers' understanding of the topic discussed across them. The full documents model is realized when readers interconnect the intertext and integrated mental models (Per-

fetti et al., 1999) by tracking who said what and by using this information to interpret and evaluate the documents' content (Britt et al., 2014).

## Previous sourcing interventions

In recent years, interventions to improve students' sourcing skills have been conducted at different educational levels (see reviews by Brand-Gruwel & van Strien, 2018; Brante & Strømsø, 2018). Modeling effective strategies, use of worksheets, prompts, guided practice, and group discussions have been common instructional methods in most of these interventions (see also Hämäläinen et al., 2020; McGrew & Byrne, 2020). Further, during interventions, students have been tasked to read multiple documents including controversies (see Brante & Strømsø, 2018; Bråten et al., 2019). At the lower educational levels, identification of source information and credibility evaluation have been emphasized whereas older students have been taught to cite sources more precisely and use source features in interpreting documents' content (see Brante & Strømsø, 2018). Even though some of the longer teacher-led interventions (e.g., Argelagós & Pifarre, 2012; Kingsley et al., 2015) conducted in the Internet context have covered the whole process of online inquiry (defining questions, searching, evaluating, synthesizing, and presenting information), sourcing has not been taught for students when specifying their information need or formulating search queries.

Next, we present three intervention studies carried out at the upper secondary school level that have aimed directly at improving students' sourcing skills. Thus, these interventions informed the ways sourcing was taught in the present study even though they were not conducted using an authentic Internet context.

Britt and Aglinskas (2002) conducted one of the first studies, comprising three short interventions (2×40 min), focused directly on students' sourcing skills. They designed a computer-based environment that prompted high school students to identify and attend to source features in history texts. The environment was designed based on principles of teaching through situated problem solving, supporting expert representations, decomposing the task, supporting transfer, providing explicit instruction, and motivating engagement. The efficacy of the interventions was tested with a sourcing test in which students read excerpts from six authentic texts that addressed controversial historical topics. While reading, they were allowed to make notes on the texts that they could later use when answering questions on the identification and evaluation of the sources and the central narrative, perspective on the controversial issue, and arguments used in the texts. For sourcing scores, correct information about the sources in students' note sheets was also counted. In all three interventions, the intervention group showed greater improvement in their scores than the controls. When computer-based and textbook-based teaching were compared, the essays produced by the group using a specially designed computer-based environment contained more source information and citations of sources than the essays of the textbook-based group.

Similarly, Braasch et al. (2013) examined the efficacy of a short (60 min) researcher-led sourcing intervention among upper secondary school students ( $N=130$ ). The intervention used a contrasting cases approach where two hypothetical adolescents, one with less and one with more sophisticated strategies, evaluated excerpts of online texts on the health risks of cell phone use. After familiarizing themselves with the cases, students were prompted to independently identify, compare, and contrast the strategies used by the hypothetical adoles-

cents. They then discussed with a partner the strategies they had identified to decide which of these were the best and why. Finally, the best strategies were collected and shared in a whole-class session. Students who participated in the intervention included more scientific concepts related to El Niño in their essays, displayed better rankings of the usefulness of the texts, gave more source-based justifications for their rankings, and more often attributed the trustworthiness of the texts to source features than those of controls.

Bråten et al. (2019) recently conducted a comprehensive sourcing intervention in natural sciences among upper secondary school students ( $N=250$ ). Compared to the studies described above, the intervention was teacher-led and markedly longer ( $9 \times 90$  min). In the scripted lessons ( $3 \times 90$  min), teachers used a contrasting cases approach (see also Braasch et al., 2013) and texts that varied in their source information. After these lessons, the students practiced the principles of adaptive sourcing through an individual writing assignment ( $3 \times 90$  min) and a group-based oral assignment ( $3 \times 90$  min). Students' performance was measured by immediate and delayed post-tests. In both tests, the students in the intervention group produced more source-based justifications for their text selections than controls. They also spent more time reading the selected texts and revisited the texts more often than controls. Further, students who participated in the intervention included more references to source features in their written products than controls.

The sourcing interventions described above have led to important understandings of how to teach sourcing skills for upper secondary school students, and younger and older students as well. For example, task assignments and reading materials applied in the lessons have included controversies related to the investigated topic (Britt & Aglinskas, 2002; Bråten et al., 2019) and/or contrasting cases approach (Braasch et al., 2013; Bråten et al., 2019) which both elicit students' sourcing behavior when reading multiple documents. Further, interventions have highlighted explicit instruction of sourcing strategies as well as students' guided practice after whole-class instruction. In addition, prompts or questions in the worksheets have been applied to enable students' independent work and to guide their attention to the specific source features at the time. In two of the studies (Braasch et al., 2013; Bråten et al., 2019), discussions with peers and in the whole class were seen as important in sharing students' ideas and learning. During the last lesson of the study by Bråten et al. (2019), students gave presentations in small groups by drawing on sources they had selected and reflecting their sourcing activities during the task. Informed by previous studies, we applied several instructional methods in designing the intervention to promote students' sourcing throughout online inquiry, such as structuring the online inquiry task, using contrasting topics and task prompts, explicit teaching of sourcing strategies, and collaborative work (see Method: Design and implementation of the intervention).

## The present study

The present study investigates the efficacy of a teacher-led intervention that aimed at enhancing upper secondary school students' sourcing during online inquiry. The design of the intervention followed the online inquiry phases (Leu et al., 2019). To facilitate students' sourcing during different phases of online inquiry and build a coherent representation of the examined issue (Perfetti et al., 1999), we applied instructional methods that have been used in previous sourcing interventions. During the intervention ( $4 \times 75$  min), students worked

collaboratively to solve a controversial health-related problem with authentic online information. Students' work was supported with explicit instruction and a joint, digital working document, including task prompts. Students' learning of sourcing skills was compared to that of control students by using a quasi-experimental pre-post design.

The following research questions were set:

RQ1. Did upper secondary school students' sourcing in different phases of an online inquiry through a teacher-led intervention increase compared to controls?

RQ2. How did students' sourcing performance change during the intervention?

RQ3. How were students' pre-intervention sourcing skills, reading fluency, prior topic knowledge, and topic order in the tasks associated with changes in their sourcing performance during the intervention?

In terms of RQ1, we assumed that the intervention group would outperform the control group in sourcing in credibility judgments and written products when their pre-sourcing skills, reading fluency, prior topic knowledge, and topic order were controlled for. The assumptions are in line with previous sourcing interventions that have successfully enhanced upper secondary school students' sourcing in their credibility evaluations, such as source-based justifications for their text selections (Bråten et al., 2019) and usefulness rankings (Braasch et al., 2013). Further, it could be assumed that students will integrate more sources to their essays after the intervention (Bråten et al., 2019; Britt & Aglinskias, 2002). Because previous interventions have not examined sourcing in specifying information need or in search querying, we did not set specific hypotheses on these sourcing practices.

In our analysis (RQ1), we controlled for students' pre-sourcing skills, their prior topic knowledge, reading fluency, and topic order. Students' pre-sourcing skills were controlled for because they are important predictors of their post-intervention performance (e.g., Hämäläinen et al., 2020; McGrew & Byrne, 2020). Prior topic knowledge and reading fluency were controlled for because of their fundamental role in reading comprehension. The reading comprehension models accentuate the role of prior knowledge when readers make meaning from the texts (Cervetti & Wright, 2020), whereas the lower-level reading skills, such as reading fluency, serve as a foundation for reading comprehension (Duke & Cartwright, 2021). Accordingly, the recent review by Anmarkrud et al. (2021) shows that the most examined cognitive skills in relation to sourcing are prior knowledge (e.g., Mason et al., 2014; Stang-Lund et al., 2019) and reading skills (e.g., Macedo-Rouet et al., 2020; Potocki et al., 2020), even though the results have been somewhat mixed.

In addition, the topic order of the texts was controlled for (RQ1) because investigated topics may elicit students' sourcing differently (Bråten et al., 2018b). For example, students have valued author expertise to a greater extent when the topic has been less familiar to them (e.g., Bråten et al., 2018b; McCrudden et al., 2016). It also seems that the relationship between individual differences and sourcing may vary with the topic addressed in reading materials (Anmarkrud et al., 2021).

In terms of RQ2, we assumed that students would differ in how their sourcing performance changed during the intervention. We expected that the substantial portion of the students, but not all, would improve their sourcing performance. For example, McGrew and Byrne (2020) conducted a sourcing intervention study among high school students, and observed students who increased, did not change, or decreased their sourcing on the online content evaluation task. Finally, we did not set any hypothesis about RQ3, as previous studies have not investigated how the above-introduced factors (pre-intervention sourcing skills,

reading fluency, prior topic knowledge, and topic order) are associated with changes in students' sourcing performance during the intervention. As these factors are related to multiple document literacy and sourcing (see Anmarkrud et al., 2021; Bråten et al., 2018c), their associations with changes in students' sourcing performance were worth solving in the present study.

## Method

### Participants

Participants comprised 365 students ( $M_{\text{age}} = 17.35$ ;  $SD = 0.40$ ) from eight upper secondary schools in Finland. Females accounted for 58.6%, which is equivalent to the proportion of females graduating from upper secondary school in Finland (Suomen virallinen tilasto [Official Statistics of Finland], 2020). In terms of parental education, 75.2% of students' mothers and 66.1% of their fathers had a tertiary level degree. Data were collected in 2018–19, before the COVID19 pandemic, during an obligatory language arts course "Texts and influence". While all students completed the tests and tasks, only the responses of those who gave their informed consent were used in this study. If a student was underage, consent was also requested from his/her guardian(s).

### Research design

We applied a quasi-experimental pre-post design with a nonequivalent control group (see Handley et al., 2018). For practical reasons, the intervention group teachers ( $N = 5$ ) were recruited based on their opportunity and willingness to implement the intervention lessons. The control group teachers ( $N = 6$ ) were not from the same schools as the intervention group teachers and were recruited after the intervention group teachers. The intervention group comprised 196 students (56.1% females) in nine courses and the control group of 169 students (61.5% females) in seven courses.

As pre- and post-tests, the students performed an online inquiry task. We counterbalanced the topic order (vaccination and fats) in both conditions. Between the tests, the intervention group participated in a teacher-led intervention ( $4 \times 75$  min lessons) on online inquiry as a part of their Texts and influence course (total of  $23 \times 75$  min lessons) while the control group participated in a regular Texts and influence course. The control group teachers received intervention materials after the completion of the study. Thus, during the study, the control group was not exposed to any of the teaching materials used in the intervention.

### Design and implementation of the intervention

To promote students' sourcing during online inquiry, we designed a teacher-led intervention that was informed by several instructional principles (see also Kiili et al., 2022). First, we designed an online inquiry task that was structured into manageable sequences (Van Merriënboer & Kirschner, 2007) following the phases of online inquiry (Leu et al., 2019) and related learning objectives (see Table 1). It is notable that for practical reasons, we were able to design a  $4 \times 75$  min unit. As a consequence, we combined the instruction of the first two



**Table 1** Phases of Online Inquiry, Learning Objectives, Description of the Sub-Tasks, and Evaluation Criteria for Intervention Lessons (4×75 min)

Lessons for online inquiry	Learning objectives	Description of the sub-tasks *	Evaluation criteria **
Lesson 1: Task assignment Planning search Locating information	Students are able to specify their information need. Students are able to select purposeful search strategies. Students are able to formulate search queries by utilizing core concepts and source information.	In your small group, select one of the four controversial health topics. Explore on the Internet what kinds of stakeholders write about the issue. Select two different stakeholders whose views you will examine more closely. Select two online texts that represent each of the stakeholders (total four texts).	Students' search plan includes main concepts about the investigated topic and related authors and venues. Students have specified purposeful search queries related to their topic and related venues. Students have selected two stakeholders differing in e.g., expertise, motives and point of views to the topic. Students have selected online texts that are suitable for the task.
Lesson 2: Evaluating information	Students are able to evaluate multiple aspects of online texts. Students are able to identify source features and evaluate them when interpreting the quality of content.	Evaluate and analyze the selected four online texts.	Students have recognized different source features and realized how those features affect the credibility and plausibility of online texts. Students have recognized the main claim in each online text and considered how well it is justified in the text. Students have utilized their notions about online text's source features when evaluating the credibility and plausibility of text's content.
Lesson 3: Synthesizing information from multiple online texts	Students are able to compose a text that compares the different views of sources and motives and evidence these views are based on. Students are able to cite the sources by providing the reader with an adequate amount of source information.	Compare the views of the stakeholders, consider potential reasons for their different views, and consider whose views are the most plausible.	Students' synthesis includes insightful considerations of similarities and differences in selected online texts (not just listed). Students have realized why critical reading on the internet is important and what kind of online texts should be relied on when making important decisions.
Lesson 4: Communicating to others the results of the inquiry	Students are able to communicate the main findings of the inquiry to other students and engage in discussions about the findings.	Present your findings to other small groups in the concluding seminar. Discuss what you have learned about critical reading online.	

\* Students' working document including prompts for each lesson can be found as [Appendix 1](#), see also Kiili et al. (2022).

\*\* Evaluation criteria were given for students before they engaged in the online inquiry task.

phases of online inquiry, i.e., specifying the information need and searching for information, into the first lesson. More emphasis was put on searching for information than on specifying

information need, thus, searching for information was taught explicitly whereas specifying the information need was taught only implicitly.

Second, we created task scenarios on controversial topics that required students to search for and select sources with different perspectives and to compare and contrast the views of the sources. Texts with contrasting views have been shown to elicit sourcing (Brante & Strømsø, 2018). Third, we designed instructional materials for teachers that they utilized in teaching explicitly effective sourcing practices that students then practiced with an online inquiry task. Explicit teaching combined with practice has shown to be an effective instructional method (Heijltjes et al., 2014). Fourth, students' sourcing was supported with a working document (Appendix 1) that included task prompts that were designed to elevate sourcing (see Gerjets et al., 2011; Kammerer et al., 2016). We also provided prompts to foster students' reflection. Finally, students' learning was supported by collaborative work (Chen et al., 2018). We created an online workspace (OneNote, Google Docs) to enable sharing and co-authoring as well as easy access to all instructional materials.

## Task

Students were tasked to explore in small groups one of four controversial health topics (cell phone radiation, food additives, the sun and health, or sleeping pills) during the four lessons of the intervention. We selected controversial topics because contradictory information seems to enhance students' attention and comparison of texts' source features (e.g., Stadler & Bromme, 2014). We also ensured beforehand that different ideas on the topic were expressed by different stakeholders on the Internet. The students were provided with four different task scenarios from which they selected one to work with in small groups. The extract below presents one of the task scenarios.

*I am a 23-year-old student from Lahti. During the last semester, I was very busy with my studies, and the situation led to sleeping difficulties. I woke up early in the mornings and couldn't sleep anymore. I visited a doctor who gave me a prescription for sleeping pills. However, my fellow student said that it can be harmful to take sleeping pills. Thus, for now, I have decided not to take the prescribed pills. Could you clarify what the Internet says about the issue?*

To orientate the students to the overall task, we provided them with a task overview that explained what they were expected to do during the four lessons. They were asked to consider the stakeholders (e.g., researchers, experts, politicians, laypersons, vendors) who were writing about the topic. They were also asked to think about why the different stakeholders were writing about the topic, the stakeholders' expertise on the issue, and the kind of evidence the stakeholders relied on in their writings. The students were also informed that they would be asked to compare the different stakeholders' points of view (e.g., commonalities, differences, tensions in points of view).

## Materials

Immediately after the pre-test and before the first intervention lesson, the students in the intervention group received an information package including the task assignment, descrip-

tion of task phases, task scenarios of alternative topics, learning objectives, and evaluation criteria (see Table 1). Analysis and reflection prompts, designed to direct students' joint work and thinking during the online inquiry, were included in the working documents (Appendix 1).

For teachers, we created a manual that included the task assignment, flow of the intervention, learning objectives, evaluation criteria, and a timetable for each lesson with links to the instructional materials. The instructional materials included information about effective online inquiry strategies, including declarative (what) and procedural (how) knowledge about the strategies and reasons why these strategies are useful. Teachers were also provided with slides that included instructions for students' working.

All the materials for students were shared digitally through Microsoft OneNote workspace. The analysis and reflection prompts and instructional materials are described in more detail in Kiili et al. (2022).

## Lessons

As described in Table 1, the four lessons followed the five phases of online inquiry: defining information need, searching for information, evaluating information, synthesizing information, and communicating the results of the online inquiry to others (Leu et al., 2019). Table 1 also describes the tasks prompted during each phase while explicit task prompts for each lesson are presented in students' working documents (Google Docs file for each group, see Appendix 1). The first three lessons were based on the teacher's instructions on sourcing in the target phase of online inquiry followed by students' group work with the Google Docs document. The teachers demonstrated the use of effective online inquiry strategies and discussed these with the students. After each lesson, the groups answered self-evaluation questions about their working and learning. The fourth lesson, a seminar, concluded the project.

The first lesson began with a teacher-led orientation to the task and students' selection of topics and groups (2–4 students). After orientation, the teacher introduced a set of effective search strategies, along with examples of how to use source information in search queries. The students then planned their information search in groups by considering and noting potential and diverse search terms in the working documents. Next, they conducted a search on the Internet and developed their search terms based on their search results. The students were then tasked to select four online texts representing two different stakeholders with different views on the topic. If needed, the selection of the online texts was completed as homework.

In the second lesson, the teacher began with an introduction to the critical evaluation of online information. For example, the teacher demonstrated how relying on only one feature of the source can lead to incorrect conclusions about the overall credibility of the text. In the following group work, students evaluated each selected online text (four texts in total) with prompts contained in their working document. They evaluated the author's/venue's expertise and intentions and considered how these were reflected in the authors' argumentation. If needed, students continued their work at home.

In the third lesson, the teacher introduced the synthesizing of information from multiple online texts and demonstrated how to connect ideas to their sources and how to provide rich information about the sources in writing. The students then practiced synthesizing by responding to the prompts in their working document. The prompts guided students to con-

sider differences and similarities in the online texts and the reasons for the differences (e.g., source features such as author's/venue's expertise and intentions). Students were also tasked to justify which of the two stakeholders' views was more plausible and note anything interesting or surprising that they had found when comparing the texts. As homework, students prepared their presentations for the seminar session.

In the seminar (fourth lesson), the teachers divided the students into groups so that the different task topics were represented, and the students selected a chair to lead each seminar group. The groups shared and discussed their main findings based on their responses recorded in the working documents. At the end of the lesson, the students self-evaluated their group work and learning during the intervention.

## Fidelity of the intervention

We ensured fidelity before and during the intervention (see McKenna et al., 2014). Before the intervention, the intervention group teachers, with one exception, participated in a three-hour-long professional development session on online inquiry. In the session, we introduced the teachers to the intervention plan, and they had an opportunity to suggest modifications. A few weeks before the intervention, we shared the revised intervention plan and intervention materials digitally with the teachers. We also assigned the teachers a researcher they could contact if they had any further questions about the lessons.

During the intervention, the teachers recorded in a diary any deviations from the intervention plan. After each lesson teachers responded to a three-point scale: The lesson was implemented 1=completely according to the plan, 2=almost according to the plan, 3=not according to the plan. Further, they were asked to write down the possible deviations from the plan. The teachers reported that the first three lessons were implemented completely or almost according to the plan ( $M=1.44$  and  $SD=0.53$  for all three lessons). The minor deviations regarded e.g., roles of absent students and time allocated for some smaller tasks. Further, for practical reasons (e.g., available space, size of group) teachers organized the fourth lesson's seminar in slightly different ways ( $M=2.22$ ,  $SD=0.67$ ).

Further, researchers observed all four lessons of three intervention group courses given by three different teachers. After the intervention, all intervention group teachers were interviewed. In addition, we collected the students' working documents before the post-tests. Observations, interviews, diaries, and completed working documents all revealed that the intervention lessons had mostly been conducted as planned.

Furthermore, we asked the control group teachers to report how much teaching they gave on online inquiry skills, as the mandatory "Texts and influence" course shared some similar learning content with the intervention (Opetushallitus, 2015). The control group teachers answered a 12-item questionnaire including four items for teaching information search, evaluation, and composing a synthesis, on a 3-point scale (1=not at all, 2=to some extent, 3=a lot). The results indicated that the control group teachers did not teach these issues very frequently in their course (means ranged as follows: 1.00–1.29 for information search, 2.00–2.29 for evaluation, and 1.29–1.57 for synthesis).

## Measures

*Reading fluency* was measured with a timed word chain test (Holopainen et al., 2004) just before the pre-test. The test consisted of 25 chains, each comprising four words written with no spaces in between. Students were asked to separate as many chains into primary words as possible in 90 seconds. The number of correctly separated words formed the total score (0–100). The test-retest reliability coefficient for the original test has varied between 0.70 and 0.84 (Holopainen et al., 2004).

*Prior topic knowledge* was measured just before the pre- and post-tests with ten statements, three correct and seven incorrect, about either vaccination or fats. Students were tasked to select the three statements they assumed to be the correct ones. They scored one point if they selected the correct statement or did not select an incorrect statement (0–1 point per statement). Four items on each topic were excluded because they were either too easy or the responses were inconsistent in relation to the responses to the remaining six items. Therefore, for each topic the score used was 0–6 points. Reliability for vaccination was 0.82 with 95% *CI* [0.68–0.96] and for fats 0.94 with 95% *CI* [0.91–0.96] (Raykov et al., 2010).

*Pre-test on sourcing / Post-test on sourcing.* We investigated students' sourcing in the pre- and post-tests by applying online inquiry assessment tasks and scoring rubrics developed in a recent study (Kiili et al., 2021). The specially designed web-based environment included instructions, task prompts, and a Google custom search engine. The students' task was to solve a health-related information problem concerning either vaccination or saturated fats. The Google custom search engine consisted of 35 preselected authentic online texts per topic, which varied in their usefulness including dimensions of source credibility and text relevance (see McCrudden, 2018). Accordingly, both topics included the same number of more useful, useful, less useful, and not useful texts (in more detail, see Hämäläinen et al., 2021).

In the task scenario of the vaccination topic, a fictitious expectant mother asked students to help her in deciding whether to vaccinate her unborn child. She had received conflicting information from two sources: a public lecture given by a civic organization that opposed vaccination and a maternity clinic nurse who favored vaccination. In turn, in the task scenario of the fats topic, a fictitious university student asked students to help him decide whether to avoid saturated fats in his diet. He had visited a book launch that took a positive stance on saturated fats and received advice from a health nurse who took the opposite view.

The task included the four phases of online inquiry (Leu et al., 2019): students (1) defined their information need; (2) searched for and selected three online texts; (3) identified and noted the main ideas in each selected text and evaluated its credibility; and (4) gave their recommendation on the issue and supported it with justifications.

As Table 2 shows, we formed four sourcing variables (Sourcing in specifying information need, Sourcing in search queries, Sourcing in credibility judgments, and Sourcing in written product) based on students' responses in the task phases. Table 2 presents the task prompts, scoring criteria, and inter-rater reliability of our scoring (Kappa) for each sourcing variable. Scoring criteria were informed by the Documents model framework (Perfetti et al., 1999; Rouet, 2006). In the first three online inquiry phases, we identified the source information that students included in their search queries and responses concerning their information need and credibility judgments. In the analysis of students' written products, we

**Table 2** Sourcing Variables in the Pre- and Post-Tests, Task Prompts in the Online Inquiry Task, Scoring Criteria and Reliability of Scoring (see Kiili et al., 2021)

Sourcing variable	Task prompts	Scoring criteria	Inter-rater reliability of scoring
Sourcing in specifying information need	<ul style="list-style-type: none"> <li>• What kind of information do you need to advice the expectant mother on whether she should vaccinate her child (vaccination topic)?</li> <li>• What kind of information do you need to advice the student on whether he should avoid saturated fats (fats topic)?</li> </ul>	0 p. = no source features or evaluative comments in the student's response 1 p. = one source feature or evaluative comment in the student's response 2 p. = two source features or/and evaluative comments in the student's response 3 p. = three or more source features or/and evaluative comments in the student's response	.76
Sourcing in search queries	<ul style="list-style-type: none"> <li>• Search for three online texts that help you to provide the expectant mother with credible information on whether she should vaccinate her child (vaccination topic).</li> <li>• Search for three on-line texts that help you to provide the student with credible information on whether he should avoid saturated fats in his diet (fats topic).</li> </ul>	Number of unique source features (organizations, credentials, names of persons relevant to the topic, type of the document) across all search queries were tallied. If student included the same source feature in multiple queries, she/he was only credited once.	.92

**Table 2** (continued)

Sourcing variable	Task prompts	Scoring criteria	Inter-rater reliability of scoring
Sourcing in credibility judgments	a) What aspects make the online text credible? b) What aspects may weaken the credibility of selected online text?	Per online text: 0 p. = no evaluation of source features (author, motivation, or venue) across the two responses (a,b). 1 p. = one source feature evaluated at least once across two responses 2 p. = two source features evaluated at least once across two responses 3 p. = all three source features evaluated at least once across two responses A sum variable with a maximum score of 9 (three online texts; 0–3 points for each) was formed. The correlations between scores of online texts varied from 0.21 to 0.41.	.75
Sourcing in written product	<ul style="list-style-type: none"> <li>• What is your position on whether the expectant mother should vaccinate her child (vaccination topic)?</li> <li>• What is your position on whether the student should avoid saturated fats (fats topic)?</li> <li>• Write below the justifications that support your position. Indicate sources that you rely on.</li> </ul>	0 p. = Student's recommendation and/or written product is NOT in line with consensus among scientists. Student's recommendation and written product is in line with consensus among scientists AND 1 p. = student does not mention any sources in his/her written product. 2 p. = student mentions specific (e.g., chemistry magazine) or implicit sources (e.g., Source 1) in his/her written product. 3 p. = student's written product includes <i>one or two indications</i> of sourcing that represent source-content link, source-source link, or evaluative statement. 4 p. = student's written product includes <i>three indications</i> of sourcing that represent source-content link, source-source link, and/or evaluative statement. 5 p. = student's written product includes <i>at least four indications</i> of sourcing that represent source-content link, source-source link, and/or evaluative statement. 6 p. = student's written product includes <i>at least five indications</i> of sourcing that represent <i>at least two categories</i> : source-content link, source-source link, and/or evaluative statement. 7 p. = student's written product includes <i>at least five indications</i> of sourcing that represent <i>all categories</i> : source-content link, source-source link, and evaluative statement.	.78

*Note.* 10% of students' responses were coded for inter-rater reliability (Kappa).

identified the source-content and source-source links and used this information in scoring the written products.

## Statistical analyses

Descriptive statistics and correlations of all employed variables are presented in [Appendix 2](#). The low pairwise correlations (max  $r = .22$ ) between predictors indicate that there is no substantial multicollinearity. In the main analyses (RQ1), the sourcing variables of the

post-test served as dependent variables and were analyzed separately. In each analysis, we controlled for the corresponding pre-test score. Group (0=control, 1=intervention) was used as the independent variable, whereas Reading fluency (0–100), Topic order (0=vaccination–fats, 1=fats–vaccination), and Prior topic knowledge (0–6) were also controlled for.

To examine the intervention effect on Sourcing in specifying information need, Sourcing in credibility judgments, and Sourcing in written product, we applied linear regression analysis. Because Sourcing in search queries was a non-normally distributed count variable with large over-dispersion, we examined its intervention effect with negative binomial regression analysis (Coxe et al., 2009).

The negative binomial regression analysis models the log of the expected count of Sourcing in search queries in the post-test (dependent variable) as a function of independent/control variables (Coxe et al., 2009). We present regression coefficients as incident rate ratios (IRRs) which were obtained by exponentiating regression coefficients using base  $e$ . For a dichotomous independent variable (i.e., Group), IRR represents the change in the expected rate of Sourcing in search queries in the post-test when the value of the independent variable changes from 0 to 1. An  $IRR > 1$  indicates how many times greater the expected rate of Sourcing in search queries in the post-test is for students in the intervention group than those in the control group. In contrast, an  $IRR < 1$  indicates that the expected rate of Sourcing in search queries in the post-test is greater for students in the control group than those in the intervention group.

With continuous control variables (i.e., Reading fluency), the IRR represents the change in the expected rate of Sourcing in search queries in the post-test when the value of the control variable increases by one unit. We determined the statistical significance of all IRRs by computing their 95% confidence intervals (CI). An IRR differs statistically significantly from the value 1 if its confidence interval does not include the value 1.

All regression analyses were conducted using Mplus statistical package (version 7.4; Muthén & Muthén, 1998–2017) with the full information maximum likelihood procedure (Enders, 2010), as missing data (0.00–0.17%) were assumed to be missing at random. Further, we estimated model parameters by using maximum likelihood estimation with non-normality robust standard errors. In the data, students were nested within 16 courses. Although intra-class correlations at the course level were small (0.01–0.11) for all variables, we used the course as a clustering variable and estimated unbiased standard errors.

Our regression analyses for RQ1 provide more general aggregate-level information on the differences between the intervention and control groups in their sourcing performance during the intervention. However, aggregate data do not necessarily apply to any specific student because the group mean may conceal individual deterioration despite improvement on average. Moreover, individual patterns of change are not revealed in the aggregate, although it is information applicable to individual students that is needed to understand who benefits from the intervention (i.e., the efficacy of the intervention).

Therefore, we supplement the analyses for RQ1 with a more individual-level examination of the effects of the intervention on students' sourcing performance (RQ2) by calculating the Reliable Change Index separately for each sourcing variable (RCI; Jacobson & Truax, 1991) for each student in the intervention group. RCI determines, for each student, if a change in the sourcing variables can be attributed to the intervention rather than chance or measurement error at  $p < .05$ , which corresponds to the value of 1.96 in the standardized normal distribution.



The RCI for an individual student was computed by dividing the difference between his/her pre- and post-test scores by the pooled standard deviation of the corresponding pre-test sourcing variable. When computing the pooled standard deviation, we used information from both the intervention and the control groups in order to take into account the potential differences between the groups in the variation. The RCI value for the individual student describes how many standard deviations his/her pre- and post-test scores differ in each sourcing variable. Next, we determined the cut-off value by counting the weighted midpoint between the pre-test means of the intervention and control groups (Atkins et al., 2005). We used individual RCI and cut-off values to classify students into those who showed a negative change during the intervention ( $\text{RCI} < -1.96$ ), those who showed no change ( $-1.96 \leq \text{RCI} \leq 1.96$ ), those who showed a reliable positive change ( $\text{RCI} > 1.96$  but did not pass the cut-off criterion), and those who also passed the cut-off criterion, thus showing a clear positive change ( $\text{RCI} > 1.96 + \text{cut-off}$ ) in their sourcing skills.

To answer RQ3, we investigated how the control variables (Pre-test scores, Reading fluency and Prior topic knowledge) were associated with the intervention group students' sourcing performance according to the RCIs. As the variable Sourcing in search queries was non-normally distributed and there were only a few students in some RCI classes, we used bootstrap analysis with 95% CIs for mean differences (Efron, 1987). When 95% CI does not include the value 0, the difference between the means of the RCI classes is statistically significant. We simulated 2 000 bootstrap samples by using bias-corrected accelerated confidence intervals (Efron, 1987) and stratified sampling according to the students' courses. Further, we investigated how topic order was associated with the intervention group students' sourcing performance according to the RCI by using crosstabulation and  $\chi^2$  test with Cramer's  $V$  for effect size.

## Results

### Descriptive statistics

Descriptive statistics of students' performance in sourcing and control variables are presented in Table 3. In the pretest, the intervention group outperformed the control group only in Sourcing in credibility judgments ( $t(342.03) = -2.05$ ,  $p = .041$ ,  $d = 0.22$ ). In all the other pre-test sourcing variables and the tests of Reading fluency and Prior topic knowledge, the intervention and the control groups performed equally, indicating no remarkable group differences at baseline.

### Efficacy of the intervention

With respect to RQ1, the regression analyses (see Table 4) showed that the intervention fostered students' attention to source features in their credibility judgments as well as their use of sources in their written products. Furthermore, the intervention group used source features in their search queries 2.23 times more often in the post-test than controls. However, the intervention did not enhance students' use of source features and evaluative statements in specifying the information need. Additionally, in the post-test, the vaccination task students performed better in all the sourcing variables than the facts task students.

**Table 3** Scores of the Sourcing and Control Variables for the Intervention and Control Groups

	Intervention group ( <i>N</i> =175–191)			Control group ( <i>N</i> =143–162)		
	<i>M</i>	<i>SD</i>	<i>Md</i>	<i>M</i>	<i>SD</i>	<i>Md</i>
Pre-test measures (observed range)						
Sourcing in specifying information need (0–3)	1.05	0.96	1	1.01	1.01	1
Sourcing in search queries (0–7)	0.37	0.87	0	0.32	0.72	0
Sourcing in credibility judgments (0–8)	3.50	1.62	3	3.16	1.47	3
Sourcing in written product (0–7)	2.96	1.93	3	2.62	1.74	3
Control variables (observed range)						
Reading fluency (4–100)	71.42	16.52	72	71.77	16.46	73
Prior topic knowledge (in the post-test) (0–6)	4.34	1.17	4	4.26	1.03	4
Post-test measures (observed range)						
Sourcing in specifying information need (0–3)	0.79	0.94	1	0.63	0.84	0
Sourcing in search queries (0–8)	0.58	1.11	0	0.24	0.54	0
Sourcing in credibility judgments (0–9)	4.41	1.77	4	3.55	1.71	4
Sourcing in written product (0–7)	3.59	2.21	4	2.68	1.88	3

**Table 4** Results of Linear ( $\beta$ ) and Negative Binomial Regression analysis (IRR; 95% CI) for the Associations Between Predictors, Independent Variable (Group) and Students' Sourcing Performance in the Post-Test

Predictors	Dependent variables			
	Post-test: Sourcing in specifying information need	Post-test: Sourcing in credibility judgments	Post-test: Sourcing in written product	Post-test: Sourcing in search queries
	$\beta$	$\beta$	$\beta$	IRR [95% CI]
Pre-test: Sourcing in specifying information need	0.44***			
Pre-test: Sourcing in credibility judgments		0.43***		
Pre-test: Sourcing in written product			0.39***	
Pre-test: Sourcing in search queries				1.32 [1.03; 1.68]*
Reading fluency	−0.02	0.10	0.08	1.02 [1.00; 1.04]
Prior topic knowledge (in the post-test)	0.02	0.03	0.10	1.28 [1.09; 1.49]*
Topic order (0=vaccination-fats, 1=fats-vaccination)	0.11**	0.21***	0.16**	2.41 [1.65; 3.53]*
<b>Group (0=control, 1=intervention)</b>	<b>0.08</b>	<b>0.19***</b>	<b>0.18**</b>	<b>2.23 [1.28; 3.88]*</b>
<b>Cohens'd for group effect [95%CI]</b>	<b>0.16 [-0.05; 0.38]</b>	<b>0.39 [0.17; 0.61]</b>	<b>0.37 [0.15; 0.58]</b>	
	$R^2 = 0.20^{***}$	$R^2 = 0.29^{***}$	$R^2 = 0.24^{***}$	

Notes:

\* The association is statistically significant when 95% CI for IRR (Incident Rate Ratio) does not include the value 1.

\*\*  $p < .01$ ; \*\*\*  $p < .001$

The rows highlighted in bold present the results between intervention group and control group with Cohen's  $d$  [95% CI] for effects.

**Table 5** Frequencies (f) and Percentages (%) of Students in the Intervention Group Demonstrating Negative Change, No Change, Reliable Positive Change and Clear Positive Change in Sourcing Variables

Sourcing variables	Negative change ( $RCI \leq -1.96$ ) f (%)	No change ( $-1.96 \leq RCI \leq 1.96$ ) f (%)	Reliable positive change ( $RCI > 1.96$ ) f (%)	Clear positive change ( $RCI > 1.96 + \text{cut-off}$ ) f (%)
Sourcing in specifying information need ( $N=172$ )	20 (11.6)	145 (84.3)	0 (0.0)	7 (4.1)
Sourcing in search queries ( $N=171$ )	31 (18.1)	98 (57.3)	0 (0.0)	42 (24.6)
Sourcing in credibility judgments ( $N=179$ )	5 (2.8)	141 (78.7)	18 (10.1)	15 (8.4)
Sourcing in written product ( $N=180$ )	14 (7.8)	129 (71.7)	20 (11.1)	17 (9.4)

Note. RCI=Reliable Change Index (Jacobson & Truax, 1991)

The RCI classes for the sourcing performance of the intervention group students are presented in Table 5. With respect to RQ2, it is notable that the number of students showing no change was high in all the sourcing variables. Further, 4.1% of the students showed a reliable or clear positive change in Sourcing in specifying information need, 24.6% in Sourcing in search queries, 18.5% in Sourcing in credibility judgments, and 20.5% in Sourcing in written product. For Sourcing in search queries, all the students demonstrating a positive change, reliable or clear, improved substantially; however, almost one-fifth of the students showed a negative change. In comparison, the changes in Sourcing in credibility judgments and written product were mostly positive.

RQ3 regarded the associations between control variables (pre-test sourcing variables, reading fluency, prior topic knowledge, and topic order) and students' RCI classes. As shown in Table 6, the intervention group students who showed a clear positive change (RCI class 4) in their sourcing performance scored the lowest in all the pre-test sourcing variables. Furthermore, the students who showed a negative change (RCI class 1) scored the highest in all the pre-test sourcing variables. Moreover, the students showing a negative change differed from the students in the other RCI classes in all the pre-test sourcing variables (see Table 7). Further, in Sourcing in credibility judgments and Sourcing in written product, the students showing a reliable change (RCI class 3) or a clear change (RCI class 4), had lower pre-test scores in corresponding sourcing variables than those showing no change (RCI class 2). In addition, students showing a clear change in Sourcing in written product had lower pre-test scores in the corresponding sourcing variable than those showing a reliable change.

With respect to the other control variables, topic order was associated with RCI classes in Sourcing in search queries ( $\chi^2(2)=15.32, p<.001, V=0.22$ ) and Sourcing in credibility judgments ( $\chi^2(3)=10.59, p=.014, V=0.18$ ). The students who explored fats in the pre-test demonstrated a clear positive change (RCI class 4) in both variables significantly more often than the students who explored vaccination in the pre-test. Conversely, the students who explored vaccination in the pre-test demonstrated a clear positive change (RCI class 4) in both variables more rarely than students who explored fats in the pre-test. Furthermore, the students who explored fats in the pre-test, demonstrated a negative change (RCI class 1) in Sourcing in search queries more rarely than the students who explored vaccination in the pre-test and vice versa. However, topic order was not associated with RCI classes for

**Table 6** Means (SD) of Intervention Group Students' RCI Classes (1=Negative Change, 2=No Change, 3=Reliable Positive Change, 4=Clear Positive Change) According to Control Variables Based on Bootstrap Analysis

Sourcing variables	RCI classes	Number of students (N)	Pre-test score <i>M</i> ( <i>SD</i> )	Reading fluency (range 0–100) <i>M</i> ( <i>SD</i> )	Prior topic knowledge (range 0–6) <i>M</i> ( <i>SD</i> )
Sourcing in specifying information need (range 0–3)	1	20	2.40 (0.50)	70.89 (15.35)	4.20 (1.06)
	2	145	0.88 (0.88)	72.68 (15.62)	4.43 (1.10)
	4	7	0.57 (0.54)	68.29 (18.20)	3.86 (2.04)
Sourcing in search queries (range 0–)	1	31	1.42 (0.72)	73.90 (13.38)	4.65 (1.05)
	2	98	0.12 (0.50)	69.21 (16.03)	4.19 (1.14)
	4	42	0.05 (0.22)	76.71 (16.16)	4.60 (1.23)
Sourcing in credibility judgments (range 0–9)	1	5	5.60 (1.14)	70.40 (14.86)	4.80 (0.84)
	2	141	3.67 (1.53)	72.09 (15.84)	4.32 (1.17)
	3	18	2.67 (1.24)	75.22 (14.25)	4.06 (1.06)
	4	15	2.27 (1.34)	68.93 (18.92)	4.67 (1.29)
Sourcing in written product (range 0–7)	1	14	5.29 (1.44)	71.43 (17.72)	4.07 (0.62)
	2	129	3.07 (1.82)	73.11 (15.97)	4.34 (1.22)
	3	20	1.95 (1.28)	69.80 (16.30)	4.35 (1.27)
	4	17	1.18 (1.02)	67.29 (12.43)	4.47 (1.01)

Note. RCI=Reliable Change Index (Jacobson & Truax, 1991). In variables Sourcing in specifying information need and Sourcing in search queries none of the students were classified in RCI class 3.

**Table 7** Comparisons of the Intervention Group Students' RCI Classes (1=Negative Change, 2=No Change, 3=Reliable Positive Change, 4=Clear Positive Change) According to Control Variables

Sourcing variables	Comparisons of RCI classes	Pre-test score	Reading fluency	Prior topic knowledge
Mean difference [95% CI] *, Cohens' <i>d</i>				
Sourcing in specifying information need ( <i>N</i> =170–172)	1 vs. 2	<b>1.52 [1.27; 1.78], <i>d</i>= -1.80</b>	-1.79 [-8.97; 5.25]	-0.23 [-0.72; 0.27]
	1 vs. 4	<b>1.83 [1.39; 2.28], <i>d</i>= -3.58</b>	2.61 [-11.94; 17.51]	0.34 [-1.06; 2.09]
	2 vs. 4	0.31 [-0.11; 0.72]	4.40 [-9.07; 18.56]	0.57 [-0.75; 2.27]
Sourcing in search queries ( <i>N</i> =170–171)	1 vs. 2	<b>1.30 [1.06; 1.57], <i>d</i>= -2.31</b>	4.70 [-0.88; 10.48]	0.46 [-0.00; 0.86]
	1 vs. 4	<b>1.37 [1.16; 1.64], <i>d</i>= -2.76</b>	-2.81 [-9.27; 4.07]	0.05 [-0.48; 0.56]
	2 vs. 4	0.07 [-0.04; 0.19]	<b>-7.51 [-13.05; -2.09], <i>d</i>=0.47</b>	-0.41 [-0.83; 0.03]
Sourcing in credibility judgments ( <i>N</i> =177–179)	1 vs. 2	<b>1.93 [0.89; 2.88], <i>d</i>= -1.27</b>	-1.69 [-21.06; 17.68]	0.48 [-0.26; 1.21]
	1 vs. 3	<b>2.93 [1.74; 4.03], <i>d</i>= -2.40</b>	-4.82 [-26.33; 16.69]	0.74 [-0.05; 1.54]
	1 vs. 4	<b>3.33 [2.07; 4.50], <i>d</i>= -2.57</b>	1.47 [-20.51; 23.44]	0.13 [-0.77; 1.10]
	2 vs. 3	<b>1.00 [0.41; 1.61], <i>d</i>= -0.67</b>	-3.13 [-13.79; 7.53]	0.26 [-0.23; 0.75]
	2 vs. 4	<b>1.40 [0.63; 2.09], <i>d</i>= -0.93</b>	3.16 [-8.40; 14.72]	-0.35 [-1.01; 0.32]
Sourcing in written product ( <i>N</i> =178–180)	3 vs. 4	0.40 [-0.47; 1.27]	6.29 [-8.59; 21.17]	-0.61 [-1.39; 0.21]
	1 vs. 2	<b>2.22 [1.38; 2.99], <i>d</i>= -1.24</b>	-1.68 [-11.10; 8.06]	-0.27 [-0.66; 0.09]
	1 vs. 3	<b>3.34 [2.39; 4.25], <i>d</i>= -2.49</b>	1.63 [-9.86; 13.40]	-0.28 [-0.89; 0.33]
	1 vs. 4	<b>4.11 [3.18; 4.89], <i>d</i>= -3.36</b>	4.13 [-6.34; 15.12]	-0.40 [-1.00; 0.14]
	2 vs. 3	<b>1.12 [0.52; 1.76], <i>d</i>= -0.64</b>	3.31 [-3.97; 10.64]	-0.01 [-0.60; 0.57]
	2 vs. 4	<b>1.89 [1.32; 2.40], <i>d</i>= -1.08</b>	5.82 [-0.66; 11.99]	-0.13 [-0.69; 0.41]
	3 vs. 4	<b>0.77 [0.10; 1.41], <i>d</i>= -0.66</b>	2.51 [-6.46; 11.35]	-0.12 [-0.84; 0.60]

Note. RCI=Reliable Change Index (Jacobson & Truax, 1991) In variables Sourcing in specifying information need and Sourcing in search queries none of the students were classified in RCI class 3.

\* The mean difference is statistically significant if 95% CI does not include 0. CI was calculated by bootstrap analysis.

Sourcing in specifying information need and Sourcing in written product. In addition to topic order, we also found an association between Reading fluency and Sourcing in search queries (see Table 7). Namely, students showing a clear change (RCI class 4) in Sourcing in search queries scored higher on Reading fluency than those showing no change (RCI class 2). Prior topic knowledge was not associated with RCI classes.

## Discussion

This study reports a sourcing intervention (4×75 min) with intervention and control groups comprising a total of over 360 upper secondary school students. Whereas previous interventions have measured students' sourcing only in one or two phases of inquiry (see Brante & Strømsø, 2018), our study focused on teaching and measuring sourcing on the Internet during the different phases of online inquiry. The uniqueness of the present study also lies in examining the characteristics of the students whose sourcing skills improved or did not improve during the intervention (cf. McGrew & Byrne, 2020). We first discuss the main findings and limitations of the study and conclude with the instructional implications of the findings.

As we expected, compared to controls, the intervention group students employed source information more often when they evaluated the credibility of online texts and composed a written product in the post-test. These results are in line with earlier findings showing that even quite short interventions can be effective in fostering upper secondary school students' sourcing skills in credibility judgments and written products (e.g., Braasch et al., 2013; Britt & Aglinskas, 2002). Further, the intervention enhanced students' use of source information when they formulated search queries.

However, sourcing in specifying information need did not increase during the intervention. This was not wholly surprising as the value of sourcing in specifying information need was not taught as explicitly as that of sourcing in the other phases of online inquiry (cf. Heijltjes et al., 2014; Marin & Halpern, 2011). This result suggests that teaching sourcing in one phase of online inquiry does not necessarily transfer to other phases of online inquiry, highlighting the importance of teaching sourcing in all the inquiry phases. Teaching why and how to source in the earlier phases of online inquiry would be important because sourcing in the earlier phases seems to support sourcing in the later phases of online inquiry (Kiili et al., 2021).

In the pre-test, students did not commonly make use of sources or source features (e.g., organizations, credentials) in their search queries. Thus, it is important to increase students' awareness and procedural knowledge about sourcing in search queries to help them broaden their strategic search repertoire. At the group level, our intervention promoted sourcing in search queries to some extent, although the students' post-test scores remained low. Notably, one-fourth of the students showed a clear positive change in their performance of sourcing in search queries. As these students had hardly engaged in sourcing when formulating search queries at the beginning of the intervention, this result suggests that they may have adopted a new sourcing practice. About one-fifth of the students performed worse in the post-test than pre-test. This may partly be explained by the topic (cf. Anmarkrud et al., 2021; Bråten et al., 2018b). It seems that it was easier to locate useful online texts on the fats topic (see Hämäläinen et al., 2021) and this did not require the students to add source information

in their queries. In sum, our study extends our understanding of sourcing during information search (see also Kiili et al., 2021) as most of the previous studies have focused on students' search strategies and reformulation of queries without paying specific attention to the use of sources in search queries (e.g., Wildemuth et al., 2018).

When prompted to evaluate the credibility of online texts, the intervention group students attended to and evaluated source features more often than controls. Likewise, the interventions by Braasch et al. (2013) and Bråten et al. (2019) enhanced upper secondary school students' use of source-based justifications for their text selections or rankings. When we examined changes in students' sourcing in credibility judgments, we found that almost one-fifth of the students showed improved performance, whereas the remainder (79%) showed no change. The students who improved had performed rather poorly in the pre-test, attending, on average, to only one source feature per online text. The intervention helped them to move towards more versatile sourcing when judging the credibility of online texts. Interestingly, the students showing no change did not perform particularly well in the pre-test either, indicating that there was no ceiling effect. These results suggest that to enhance students' critical online reading skills, there is a need to regularly teach sourcing when students read online texts varying in quality.

In addition, it seems that different texts elicit different kinds of sourcing behavior (cf. Bråten et al., 2011; 2015). For example, in the present study, some authentic online texts missed the name of the author and in some texts, the author's motives were more obvious than in others. Even though students responded to the separate questions regarding aspects that strengthened and aspects that weakened the credibility of online texts, they did not attend to and evaluate consistently source features (author, venue, intentions) through different texts, not even in the post-test. However, paying attention to the author expertise should be regularly used sourcing practice (e.g., Bråten et al., 2018b).

Further, the intervention enhanced students' use of source information in their written products when justifying their stance on vaccinating a child or avoiding saturated fats (see also Bråten et al., 2019; Britt & Aglinskias, 2002). It should be noted that the scores of students' written products included the mentioned sources but also the use of evaluative statements, source-source links, and source-content links (see Perfetti et al., 1999). Again, the students with the weakest skills in the pre-test were mostly those who showed improvement (altogether 20.5% improved) in the post-test. This means that they had hardly used the links or evaluative statements in their written products before the intervention and that the intervention guided them towards the more sophisticated sourcing practices that are required to build an intertext model (see Perfetti et al., 1999).

It is notable that the students were allowed to consult their self-selected online texts when composing the written product (cf. Bråten et al., 2019), a procedure which makes this subtask easier than when based solely on memory and mental representations, as in some earlier studies (e.g., Braasch et al., 2013; Britt & Aglinskias, 2002). However, our task also resembles basic school assignments as well as expert practices, where documents are usually available when composing a written synthesis (cf. Vandermeulen et al., 2020).

Despite our expectations, the number of students whose sourcing performance improved was limited (4–25% across different sourcing practices). However, the intervention especially fostered the performance of the students with the weakest sourcing skills in the pre-test. This result is important as very limited sourcing skills may result in the recurring use of dis- and misinformation (Sinatra & Lomabardi, 2020). Thus, the students whose perfor-

mance did not change during the intervention had better sourcing skills to start with than those whose performance improved.

Some of the more advanced students also performed worse in the post-test than pre-test. This may partly be explained by the test topics, which seemed to elicit sourcing activity somewhat differently (cf. Anmarkrud et al., 2021). It is also possible that some students were not sufficiently motivated to put effort into the post-test assignment (see Bråten et al., 2018a; List & Alexander, 2018). Alternative explanations may relate to the small group work. Teachers reported variation in students' engagement, some small groups were more engaged than others. It may also well be that some groups did not have an optimal construction for learning. Accordingly, small groups including students with weaker and better skills, may serve students with better skills if they are the ones giving the elaborated help for peers with weaker skills (see review by Wilkinson & Fung, 2002).

Students' prior topic knowledge and reading fluency were not associated, with one exception, with their sourcing skills in the post-test and the changes in their sourcing performance during the intervention. The recent review by Anmarkrud et al. (2021) reported mixed results on the contribution of reading fluency and prior knowledge to students' sourcing skills. The authors suggested that mixed results may be related to the used measures (Anmarkrud et al., 2021). Our results regarding the role of prior topic knowledge are in line with the study by Kammerer et al. (2016), who likewise applied true/false items, and did not find an association between students' prior topic knowledge and their sourcing skills. Further, in our study, the prior knowledge measure only included six items. In terms of reading fluency, upper secondary school students have probably reached a reasonable level so that it does not hinder them in acquiring sourcing skills. It is notable that in Finland, after 9 years of compulsory comprehensive school, about half of the students select academic-oriented upper secondary school.

## Limitations and future research

The study also has its limitations. First, we arranged a three-hour professional development session for the teachers of the intervention group a couple of weeks before the intervention. Although this included an introduction to critical online reading skills, the time was quite short for teachers to reach a profound understanding of sourcing in online reading. In future studies, a longer and more recurrent training program (cf. Bråten et al., 2019) could better equip teachers to teach sourcing during online inquiry and also challenge the competencies of students possessing better sourcing skills.

Second, sourcing in specifying information need was not taught as explicitly during the intervention as sourcing in the other phases of online inquiry. It was only implicitly embedded in the task assignment and in the working document when students planned their information search. In the future, studies should improve the efficacy of their interventions by including more explicit teaching on sourcing when defining information need.

Third, because the content of language arts courses in upper secondary school is very broad, the teachers were not able to find more time for us to investigate the sustainability of the results with a delayed post-test. As our results showed different-level changes in students' sourcing performance during the intervention, in future research, it would be important to ascertain how permanent these changes are. It should be noted that the similarity of

the intervention group students' outcome means across the differently timed post-tests does not tell us how sustainable the learned skills are if changes at the individual level from one post-test to another are not also measured (cf. Bråten et al., 2019).

## Instructional implications

Our results suggest that the designed sourcing intervention has the potential to promote upper secondary school students' sourcing skills. This requires the explicit teaching of sourcing practices and the sequenced practicing of strategies that follow the four online inquiry phases. Our study revealed that diverging from these principles is not worthwhile. Thus, educators applying the developed intervention should ensure to explicitly teach all inquiry practices, including sourcing in specifying information need (cf. Heijltjes et al., 2014; Marin & Halpern, 2011).

The instructional methods used in this study seemed particularly beneficial for the students with the weakest sourcing skills. Thus, highlighting the attention to, evaluation, and use of source information through modeling, lecturing, and scaffolding students with guiding questions represent efficient methods of teaching sourcing (cf. Brante & Strømsø, 2018). Students with the weaker skills may also profit from discussing and exploring a controversial topic in small groups, as this provides them with opportunities to discover more ways to evaluate, use, and interpret source information in online texts (e.g., Kiili et al., 2019). Although the students in the present study were allowed to form the small groups by themselves, the scaffolded small-group work combined with explicit teaching seemed to be an efficient method for students with the weakest skills to learn sourcing skills during online inquiry (cf. Wilkinson & Fung, 2002).

Despite these promising results, our intervention did not serve as effectively the students who performed better in the pre-test than students who had the weakest skills in the pre-test. This suggests that more attention should be put to differentiating instruction, for example, by ensuring a sufficient difficulty level of the tasks. This need was supported by teachers' comments in their diaries. They reported that even though the prompts offered opportunities for students to practice sourcing at their own level, they observed that some tasks were too easy or too difficult for part of the students.

Although the present intervention was designed for upper secondary school students, teaching sourcing throughout online inquiry could be scaled down for secondary and even upper primary school students. This would require the use of more concrete concepts throughout the task. For younger students, sourcing in search queries could be limited to professions and selected texts to two contradictory ones written by a professional and a layperson. As sourcing in written texts is particularly challenging for primary and secondary school students (Kiili et al., 2020; Pérez et al., 2018), students' composition of a written product could be scaffolded with sentence starters requiring integration of sources in their writing. Whatever the means of facilitation, it is critical that also younger students also experience sourcing when engaging in online inquiry.

There are several ways how our intervention can be improved. First, providing feedback on students' sourcing during online inquiry could scaffold students towards more sophisticated sourcing practices. In the present study, our design did not include any systematic feedback procedures or guidelines for the teachers even though feedback plays a crucial role



in students' learning (Hattie & Timperley, 2007; Van der Kleij et al., 2015). The external feedback from a teacher is essential (Huisman et al., 2019), but in some circumstances, peer feedback can be as effective as teacher feedback (Huisman et al., 2019). Importantly, peer feedback not only benefits the receiver but also the provider, as it requires students to actively consider the criteria for advanced sourcing (Huisman et al., 2018) and helps them to reflect on their own sourcing skills (Van Popta et al., 2017).

Secondly, more attention could be paid to designing engaging tasks. In the present study, we designed four alternative task scenarios on health issues that were connected to young people's lives. According to teacher and student feedback, the topics did not, however, initiate interest among some students (see Kiili et al., 2022). This accentuates the importance of selecting online inquiry topics that are both topical and novel among young people (cf. Anmarkrud et al., 2021). At their best, topics will stimulate productive emotions, such as curiosity and enjoyment (Chinn et al., 2021).

## Conclusions

When reading and learning through online information, sourcing is one of the key practices supporting the evaluation of information, comprehension of multiple viewpoints, and decision-making (Scharrer & Salmerón, 2016). Sourcing is also an overarching practice that can occur throughout online inquiry, starting from the point when readers turn to the Internet to solve a problem and ending when they communicate their findings to others (Kiili et al., 2021). Our study suggests that sourcing can be taught throughout the online inquiry process by carefully designing sourcing practices as an integral part of online inquiry.

The rapid spread of false information online has increased concerns about the vulnerability of children and adolescents with low critical reading skills (Howard et al., 2021). For example, adolescents who use social media frequently tend to overlook sources' credibility (e.g., Macedo-Rouet et al., 2020) which may lead them to spread disinformation unintentionally. Encouragingly, the intervention implemented here succeeded in enhancing the sourcing skills of the students with the weakest skills. However, sourcing is not effortless for adolescents or easy to teach for them and thus, promotion of sourcing should be a continuous effort and implemented in different school subjects.

**Supplementary Information** The online version contains supplementary material available at <https://doi.org/10.1007/s11251-022-09597-2>.

**Acknowledgements** This research was funded by the Academy of Finland (Project Number: 285806). The authors would like to thank Timo Salminen and Tuulikki Alamettälä for the data collection, Jari Hämäläinen, Symcode Oy for the software development, Michael Freeman for his valuable comments on the language, and the students and teachers who participated in the study.

**Author contributions** **Elina K. Hämäläinen**: Conceptualization, Methodology, Resources, Formal analysis, Writing - original draft preparation.

**Carita Kiili** Conceptualization, Methodology, Resources, Writing – review and editing, Supervision.

**Eija Räikkönen** Methodology, Formal analysis, Writing – review and editing, Supervision.

**Minna Lakkala** Conceptualization, Resources, Investigation, Writing – review and editing.

**Liisa Ilomäki** Conceptualization, Resources, Investigation, Writing – review and editing.

**Auli Toom** Conceptualization, Writing – review and editing, Funding Acquisition.

**Miika Marttunen** Conceptualization, Writing – review and editing, Funding Acquisition, Supervision.

**Funding** Open Access funding provided by University of Jyväskylä (JYU).

**Declarations** This research was funded by the Academy of Finland (Project Number: 285,806). The authors have no potential conflicts of interest to disclose. Data and procedures have met all ethical guidelines and standards from our institutions. Informed consent was asked from all students and if a student was underage, consent was also requested from his/her guardian(s). The datasets generated during and/or analyzed during the current study are available from the corresponding author on reasonable request.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

## References

- Anmarkrud, Ø., Bråten, I., Florit, E., & Mason, L. (2021). The role of individual differences in sourcing: A systematic review. *Educational Psychology Review*, 54(2), 61–83. <https://doi.org/10.1007/s10648-021-09640-7>
- Argelagos, E., & Pifarré, M. (2012). Improving information problem solving skills in secondary education through embedded instruction. *Computers in Human Behavior*, 28(2), 515–526. <https://doi.org/10.1016/j.chb.2011.10.024>
- Atkins, D. C., Bedics, J. D., McGlinchey, J. B., & Beauchaine, T. P. (2005). Assessing clinical significance: Does it matter which method we use? *Journal of Consulting and Clinical Psychology*, 73(5), 982–989. <https://doi.org/10.1037/0022-006X.73.5.982>
- Barzilai, S., Thomm, E., & Shlomi-Elouz, 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, 101367. <https://doi.org/10.1016/j.learninstruc.2020.101367>
- Barzilai, S., Tzadok, E., & Eshet-Alkalai, Y. (2015). Sourcing while reading divergent expert accounts: Pathways from views of knowing to written argumentation. *Instructional Science*, 43(6), 737–766. <https://doi.org/10.1007/s11251-015-9359-4>
- Ben-Eliyahu, A., Moore, D., Dorph, R., & Schunn, C. D. (2018). Investigating the multidimensionality of engagement: Affective, behavioral, and cognitive engagement across science activities and contexts. *Contemporary Educational Psychology*, 53, 87–105. <https://doi.org/10.1016/j.cedpsych.2018.01.002>
- Braasch, J. L., Bråten, I., Strømsø, H. I., & Anmarkrud, Ø. (2013). Promoting secondary school students' evaluation of source features of multiple documents. *Contemporary Educational Psychology*, 38(3), 180–195. <https://doi.org/10.1016/j.cedpsych.2013.03.003>
- Brand-Gruwel, S., & van Strien, J. L. (2018). Instruction to promote information-problem solving on the Internet in primary and secondary education students: A systematic literature review. In J. L. Braasch, I. Bråten & M. T. McCrudden (Eds.), *Handbook of multiple source use* (pp. 401–422). Routledge.
- Brante, E. W., & Strømsø, H. I. (2018). Sourcing in text comprehension. A review of interventions targeting sourcing skills. *Educational Psychology Review*, 30(3), 773–799. <https://doi.org/10.1007/s10648-017-9421-7>
- Bråten, I., Braasch, J. L., & Strømsø, H. I. (2015). Establishing trustworthiness when students read multiple documents containing conflicting scientific evidence. *Reading Psychology*, 36(4), 315–349. <https://doi.org/10.1080/02702711.2013.864362>

- Bråten, I., Brante, E. W., & Strømso, H. I. (2018a). What really matters: The role of behavioural engagement in multiple document literacy tasks. *Journal of Research in Reading*, 41(4), 680–699. <https://doi.org/10.1111/1467-9817.12247>
- Bråten, I., Brante, E. W., & Strømso, H. I. (2019). Teaching sourcing in upper secondary school: A comprehensive sourcing intervention with follow-up data. *Reading Research Quarterly*, 54(4), 481–505. <https://doi.org/10.1002/rrq.253>
- Bråten, I., McCrudden, M. T., Stang Lund, E., Brante, E. W., & Strømso, H. I. (2018b). 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(3), 345–365. <https://doi.org/10.1002/rrq.197>
- Bråten, I., Stadtler, M., & Salmerón, L. (2018c). The role of sourcing in discourse comprehension. In M. F. Schober, D. N. Rapp, & Britt, M. A. (Eds.), *Handbook of discourse processes* (2nd ed, pp. 141–166). Routledge.
- Bråten, I., Strømso, H. I., & Salmerón, L. (2011). Trust and mistrust when students read multiple information sources about climate change. *Learning and Instruction*, 21(2), 180–192. <https://doi.org/10.1016/j.learninstruc.2010.02.002>
- Britt, M. A., & Aglinskias, C. (2002). Improving students' ability to identify and use source information. *Cognition and Instruction*, 20(4), 485–522. [https://doi.org/10.1207/S1532690XCI2004\\_2](https://doi.org/10.1207/S1532690XCI2004_2)
- 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. <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. 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(2), 59–69. <https://doi.org/10.1080/00461520.2014.921572>
- Cervetti, G. N., & Wright, T. S. (2020). The role of knowledge in understanding and learning from text. In E. B. Moje, P. Afflerbach, P. Enciso, & N. K. Leseaux (Eds.), *Handbook of reading research* (Vol. 5, pp. 237–260). Routledge.
- Chen, J., Wang, M., Kirschner, P. A., & Tsai, C. C. (2018). The role of collaboration, computer use, learning environments, and supporting strategies in CSCL: A meta-analysis. *Review of Educational Research*, 88, 799–843. <https://doi.org/10.3102/0034654318791584>
- Chinn, C. A., Barzilai, S., & Duncan, R. G. (2021). Education for a “post-truth” world. New directions for research and practice. *Educational Researcher*, 50(1), 51–60. <https://doi.org/10.3102/0013189X20940683>
- Coiro, J. (2021). Toward a multifaceted heuristic of digital reading to inform assessment research, practice, and policy. *Reading Research Quarterly*, 56(1), 9–31. <https://doi.org/10.1002/rrq.302>
- Coxe, S., West, S. G., & Aiken, L. S. (2009). The analysis of count data. A gentle introduction to Poisson regression and its alternatives. *Journal of Personality Assessment*, 91(2), 121–136. <https://doi.org/10.1080/00223890802634175>
- Duke, N. K., & Cartwright, K. B. (2021). The science of reading progresses: Communicating advances beyond the simple view of reading. *Reading Research Quarterly*, 56, S25–S44. <https://doi.org/10.1002/rrq.411>
- Efron, B. (1987). Better bootstrap confidence intervals. *Journal of the American Statistical Association*, 82(397), 171–185.
- Enders, C. K. (2010). *Applied missing data analysis*. Guilford Publications.
- 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(2), 220–231. <https://doi.org/10.1016/j.learninstruc.2010.02.005>
- Haas, A., & Unkel, J. (2017). Ranking versus reputation: perception and effects of search result credibility. *Behaviour & Information Technology*, 36(12), 1285–1298. <https://doi.org/10.1080/0144929X.2017.1381166>
- Hahnel, C., Eichmann, B., & Goldhammer, F. (2020). Evaluation of Online Information in University Students: Development and Scaling of the Screening Instrument EVON. *Frontiers in Psychology*, 11, 562128. <https://doi.org/10.3389/fpsyg.2020.562128>
- Handley, M. A., Lyles, C. R., McCulloch, C., & Cattamanchi, A. (2018). Selecting and improving quasi-experimental designs in effectiveness and implementation research. *Annual Review of Public Health*, 39, 5–25. <https://doi.org/10.1146/annurev-publhealth-040617-014128>
- Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of Educational Research*, 77(1), 81–112. <https://doi.org/10.1146/annurev-publhealth-040617-014128>

- 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(8), 761–773. <https://doi.org/10.1080/0144929X.2018.1477992>
- Heijltjes, A., Van Gog, T., & Paas, F. (2014). Improving students' critical thinking: Empirical support for explicit instructions combined with practice. *Applied Cognitive Psychology*, 28(4), 518–530. <https://doi.org/10.1002/acp.3025>
- Holopainen, L., Kairaluoma, L., Nevala, J., Ahonen, T., & Aro, M. (2004). *Lukivaikkeusienseulontatesti nuorille ja aikuisille [Dyslexia screening test for youth and adults]*. Niilo Mäki Institute
- Howard, P. N., Neudert, L. M., & Prakash, N. (2021). *Digital misinformation/disinformation and children*. Unicef. <https://www.unicef.org/globalinsight/media/2096/file/UNICEF-Global-Insight-Digital-Mis-Disinformation-and-Children-2021.pdf>
- Huisman, B., Saab, N., van den Broek, P., & van Driel, J. (2019). The impact of formative peer feedback on higher education students' academic writing: a Meta-Analysis. *Assessment & Evaluation in Higher Education*, 44(6), 863–880. <https://doi.org/10.1080/02602938.2018.1545896>
- Huisman, B., Saab, N., Van Driel, J., & Van Den Broek, P. (2018). Peer feedback on academic writing: undergraduate students' peer feedback role, peer feedback perceptions and essay performance. *Assessment & Evaluation in Higher Education*, 43(6), 955–968. <https://doi.org/10.1080/02602938.2018.1424318>
- 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, 106372. <https://doi.org/10.1016/j.chb.2020.106372>
- Hämäläinen, E. K., Kiili, C., Räikkönen, 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*, 37(5), 1409–1422. <https://doi.org/10.1111/jcal.12580>
- Jacobson, N. S., & Truax, P. (1991). Clinical significance: A statistical approach to defining meaningful change in psychotherapy research. *Journal of Consulting and Clinical Psychology*, 59, 12–19.
- 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. <https://doi.org/10.1016/j.compedu.2016.07.001>
- Kiili, C., Bråten, I., Kullberg, N., & Leppänen, P. H. (2020). Investigating elementary school students' text-based argumentation with multiple online information resources. *Computers & Education*, 147, 103785. <https://doi.org/10.1016/j.compedu.2019.103785>
- Kiili, C., Coiro, J., & Räikkönen, E. (2019). Students' evaluation of information during online inquiry: Working individually or in pairs. *Australian Journal of Language and Literacy*, 42(3), 167–183.
- Kiili, C., Forzani, E., Brante, E. W., Räikkönen, E., & Marttunen, M. (2021). Sourcing on the internet: Examining the relations among different phases of online inquiry. *Computers and Education Open*, 2, 100037. <https://doi.org/10.1016/j.caeo.2021.100037>
- Kiili, C., Lakkala, M., Ilomäki, L., Toom, A., Coiro, J., Hämäläinen, E., & Sormunen, E. (2022). Designing classroom practices for teaching online inquiry: Experiences from the field. *Journal of Adolescent & Adult Literacy*, 65(4), 297–308. <https://doi.org/10.1002/jaal.1206>
- Kiili, C., Leu, D. J., Utriainen, J., Coiro, J., Kanninen, L., Tolvanen, A., Lohvansuu, K., & Leppänen, P. H. (2018). Reading to learn from online information: Modeling the factor structure. *Journal of Literacy Research*, 50(3), 304–334. <https://doi.org/10.1177/1086296X18784640>
- 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. Retrieved from [https://scholarworks.wmich.edu/reading\\_horizons/vol54/iss2/5](https://scholarworks.wmich.edu/reading_horizons/vol54/iss2/5)
- Kobayashi, K. (2014). Students' consideration of source information during the reading of multiple texts and its effect on intertextual conflict resolution. *Instructional Science*, 42(2), 183–205. <https://doi.org/10.1007/s11251-013-9276-3>
- 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, & Ruddell, R. B. (Eds.). *Theoretical models and processes of literacy* (7th ed., pp. 319–346). Taylor & Francis.
- List, A., & Alexander, P. A. (2018). Cold and warm perspectives on the cognitive affective engagement model of multiple source use. In J. L. Braasch, I. Bråten, & M. T. McCrudden (Eds.), *Handbook of multiple source use* (pp. 34–54). Routledge.
- List, A., Alexander, P. A., & Stephens, L. A. (2017). Trust but verify: Examining the association between students' sourcing behaviors and ratings of text trustworthiness. *Discourse Processes*, 54(2), 83–104. <https://doi.org/10.1080/0163853X.2016.1174654>

- Macedo-Rouet, M., Salmerón, L., Ros, C., Pérez, A., Stadler, M., & Rouet, J. F. (2020). Are frequent users of social network sites good information evaluators? An investigation of adolescents' sourcing abilities (¿ Son los usuarios frecuentes de las redes sociales evaluadores competentes? Un estudio de las habilidades de los adolescentes para identificar, evaluar y hacer uso de las fuentes). *Journal for the Study of Education and Development*, 43(1), 101–138. <https://doi.org/10.1080/02103702.2019.1690849>
- Marin, L. M., & Halpern, D. F. (2011). Pedagogy for developing critical thinking in adolescents: Explicit instruction produces greatest gains. *Thinking Skills and Creativity*, 6(1), 1–13. <https://doi.org/10.1016/j.tsc.2010.08.002>
- 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. <https://doi.org/10.1016/j.compedu.2014.03.016>
- McCrudden, M. T. (2018). Text relevance and multiple-source use. In J. L. Braasch, I. Bråten, & M. T. McCrudden, (Eds.). *Handbook of multiple source use* (pp. 168–183). Routledge.
- McCrudden, M. T., Stenseth, T., Bråten, I., & Strømso, H. I. (2016). The effects of topic familiarity, author expertise, and content relevance on Norwegian students' document selection: A mixed methods study. *Journal of Educational Psychology*, 108(2), 147–162. <https://doi.org/10.1037/edu0000057>
- McGrew, S., Breakstone, J., Ortega, T., Smith, M., & Wineburg, S. (2018). Can students evaluate online sources? Learning from assessments of civic online reasoning. *Theory & Research in Social Education*, 46(2), 165–193. <https://doi.org/10.1080/00933104.2017.1416320>
- McGrew, S., & 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. <https://doi.org/10.1080/15391523.2020.1795956>
- McKenna, J. W., Flower, A., & Ciullo, S. (2014). Measuring fidelity to improve intervention effectiveness. *Intervention in School and Clinic*, 50(1), 15–21. <https://doi.org/10.1177/1053451214532348>
- Muthén, L., & Muthén, B. O. (1998–2017). *Mplus user's guide* (8th ed). Muthén & Muthén.
- Nicol, D. J., & Macfarlane-Dick, D. (2006). Formative assessment and self-regulated learning: A model and seven principles of good feedback practice. *Studies in Higher education*, 31(2), 199–218. <https://doi.org/10.1080/03075070600572090>
- Opetushallitus (2015). *Lukion opetussuunnitelman perusteet. [National core curriculum for upper secondary education]*. Opetushallitus.
- Paul, J., Cerdán, R., Rouet, J. F., & Stadler, M. (2018). Exploring fourth graders' sourcing skills / Un análisis de la capacidad de escrutinio sobre las fuentes de información de los estudiantes de cuarto grado. *Infancia y Aprendizaje*, 41(3), 536–580. <https://doi.org/10.1080/02103702.2018.1480458>
- Pérez, A., Potocki, A., Stadler, M., Macedo-Rouet, M., Paul, J., Salmerón, L., & Rouet, J. F. (2018). Fostering teenagers' assessment of information reliability: Effects of a classroom intervention focused on critical source dimensions. *Learning and Instruction*, 58, 53–64. <https://doi.org/10.1016/j.learninstruc.2018.04.006>
- 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(1), 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(2), 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(2), 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*. Psychology Press.
- Salmerón, L., Gil, L., & Bråten, I. (2018). Effects of reading real versus print-out versions of multiple documents on students' sourcing and integrated understanding. *Contemporary Educational Psychology*, 52, 25–35. <https://doi.org/10.1016/j.cedpsych.2017.12.002>
- Scharrer, L., & Salmerón, L. (2016). Sourcing in the reading process. Introduction to the special issue. *Reading and Writing*, 29(8), 1539–1548. <https://doi.org/10.1007/s11145-016-9676-2>
- 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(3), 120–131. <https://doi.org/10.1080/00461520.2020.1730181>

- Stadtler, M., & Bromme, R. (2014). The content-source integration model: A taxonomic description of how readers comprehend conflicting scientific information. In D. N. Rapp, & J. L. G. Braasch, (Eds.). *Processing inaccurate information: Theoretical and applied perspectives from cognitive science and the educational sciences* (pp. 379–402). MIT Press.
- Stang Lund, E., Bråten, I., Brandmo, C., Brante, E. W., & Strømsø, H. I. (2019). Direct and indirect effects of textual and individual factors on source-content integration when reading about a socio-scientific issue. *Reading and Writing*, 32(2), 335–356. <https://doi.org/10.1007/s11145-018-9868-z>
- Strømsø, H. I., & Bråten, I. (2014). Students' sourcing while reading and writing from multiple web documents. *Nordic Journal of Digital Literacy*, 9(02), 92–111. <https://doi.org/10.18261/ISSN1891-943X-2014-02-02>
- Strømsø, H. I., Bråten, I., Britt, M. A., & Ferguson, L. E. (2013). Spontaneous sourcing among students reading multiple documents. *Cognition and Instruction*, 31(2), 176–203. <https://doi.org/10.1080/07370008.2013.769994>
- Suomen virallinen tilasto [Official Statistics of Finland], (2020). *Opiskelijat ja tutkinnot [Students and degrees]*. Liitetaulukko 1. Lukiokoulutuksen opiskelijat ja ylioppilastutkinnon suorittaneet 2001–2020. [Appendix 1. Students who graduated from upper secondary school and completed the matriculation examination 2001–2020.] Tilastokeskus. [http://www.stat.fi/til/opiskkt/2020/opiskkt\\_2020\\_2021-06-10\\_tau\\_001\\_fi.html](http://www.stat.fi/til/opiskkt/2020/opiskkt_2020_2021-06-10_tau_001_fi.html)
- Tierney, R. J., & Pearson, P. D. (2021). *A history of literacy education: Waves of research and practice*. Teachers College Press.
- Van der Kleij, F. M., Feskens, R. C., & Eggen, T. J. (2015). Effects of feedback in a computer-based learning environment on students' learning outcomes: A meta-analysis. *Review of Educational Research*, 85(4), 475–511. <https://doi.org/10.3102/0034654314564881>
- Vandermeulen, N., Van den Broek, B., Van Steendam, E., & Rijlaarsdam, G. (2020). In search of an effective source use pattern for writing argumentative and informative synthesis texts. *Reading and Writing*, 33(2), 239–266. <https://doi.org/10.1007/s11145-019-09958-3>
- Van Merriënboer, J. J., & Kirschner, P. A. (2007). *Ten steps to complex learning: A systematic approach to four-component instructional design*. Mahwah, NJ: Lawrence Erlbaum.
- Van Popta, E., Kral, M., Camp, G., Martens, R. L., & Simons, P. R. (2017). Exploring the value of peer feedback in online learning for the provider. *Educational Research Review*, 20, 24–34. <https://doi.org/10.1016/j.edurev.2016.10.003>
- Wildemuth, B. M., Kelly, D., Boettcher, E., Moore, E., & Dimitrova, G. (2018). Examining the impact of domain and cognitive complexity on query formulation and reformulation. *Information Processing & Management*, 54(3), 433–450. <https://doi.org/10.1016/j.ipm.2018.01.009>
- Wilkinson, I. A., & Fung, I. Y. (2002). Small-group composition and peer effects. *International Journal of Educational Research*, 37(5), 425–447. [https://doi.org/10.1016/S0883-0355\(03\)00014-4](https://doi.org/10.1016/S0883-0355(03)00014-4)
- 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>