JYU DISSERTATIONS 663

Elina Hämäläinen

Examining and Enhancing Adolescents' Critical Online Reading Skills

UNIVERSITY OF JYVÄSKYLÄ
FACULTY OF EDUCATION AND PSYCHOLOGY

JYU DISSERTATIONS 663

Elina Hämäläinen

Examining and Enhancing Adolescents' Critical Online Reading Skills

Esitetään Jyväskylän yliopiston kasvatustieteiden ja psykologian tiedekunnan suostumuksella
julkisesti tarkastettavaksi Agoran Lea Pulkkisen salissa
syyskuun 2. päivänä 2023 kello 12.

Academic dissertation to be publicly discussed, by permission of
the Faculty of Education and Psychology of the University of Jyväskylä,
in building Agora, Lea Pulkkinen Hall, on September 2, 2023, at 12 o'clock.

JYVÄSKYLÄN YLIOPISTO
UNIVERSITY OF JYVÄSKYLÄ

JYVÄSKYLÄ 2023

Editors
Miika Marttunen
Department of Education, University of Jyväskylä
Päivi Vuorio
Open Science Centre, University of Jyväskylä

ISBN 978-951-39-9654-3 (PDF)
URN:ISBN:978-951-39-9654-3
ISSN 2489-9003

Permanent link to this publication: http://urn.fi/URN:ISBN:978-951-39-9654-3

ABSTRACT

Hämäläinen, Elina
Examining and Enhancing Adolescents' Critical Online Reading Skills
Jyväskylä: University of Jyväskylä, 2023, 93 p. + original papers
(JYU Dissertations
ISSN 2489-9003; 663)
ISBN 978-951-39-9654-3 (PDF)

This dissertation examined whether interventions as part of regular schoolwork could enhance adolescents' critical online reading skills. Further, students' critical online reading skills were explored and investigated how different factors were related to students' skills and the changes in their skills during intervention. In the pre- and post-tests, students completed an online inquiry task, including phases of searching, selecting, evaluating, and synthesizing information.

In Sub-study I was mainly examined whether an intervention (21 × 45 min lessons) affected sixth graders' (N = 342) justifications for the credibility of online texts. The intervention comprised the explicit teaching of online inquiry skills and the practice of these skills in two projects. The results showed that justifying the credibility of the online texts was difficult for most sixth graders. However, after the intervention, students evaluated the source information of online texts (e.g., author and venue) more often than the control group.

In Sub-study II it was explored how upper secondary school students (N = 372) justified the credibility of online texts and whether their Internet-specific epistemic justifications were related to their evaluation performance. The results revealed considerable discrepancies in students' abilities according to different credibility aspects (author, venue, intentions, evidence, and corroboration) and the depth of their reasoning. The students who selected more useful online texts and believed that they evaluated the authority or compared multiple texts when reading online were better at justifying the credibility of online texts.

In Sub-study III it was examined whether an intervention (4 × 75 min) increased upper secondary school students' (N = 365) sourcing skills during online inquiry. Students investigated their topic in small groups according to the questions and tasks in the working document. The teacher's short introductions on online inquiry skills supported students' work. Students increased their sourcing performance in search queries, credibility judgments, and written products compared with the control group. Furthermore, students with the weakest skills benefited the most from the intervention.

In all Sub-studies I–III, students' basic reading skills and the topic of the online inquiry task were associated with their critical online reading skills or the changes in their skills during the intervention.

Keywords: critical reading, credibility evaluation, sourcing, online inquiry, intervention, primary school, upper secondary school

TIIVISTELMÄ (ABSTRACT IN FINNISH)

Hämäläinen, Elina
Nuorten kriittisen nettilukemisen taitoja tutkimassa ja tukemassa
Jyväskylä: Jyväskylän yliopisto, 2023, 93 s. + alkuperäiset julkaisut
(JYU Dissertations
ISSN 2489-9003; 663)
ISBN 978-951-39-9654-3 (PDF)

Tässä väitöstutkimuksessa kehitettiin opetusmenetelmiä nuorten kriittisen nettilukemisen taitojen tukemiseen. Interventioiden vaikuttavuutta selvitettiin tutkivan nettilukemisen tehtävällä, jossa nuoret etsivät, valitsivat ja arvioivat nettitekstejä sekä laativat kirjoitelman nettitekstien pohjalta. Lisäksi tutkittiin nuorten kriittisen nettilukemisen taitoja ja niihin yhteydessä olevia tekijöitä.

Osatutkimuksessa I tarkasteltiin, edistikö tutkivan nettilukemisen interventio (21 x 45 min) kuudesluokkalaisten (N = 342) taitoja perustella nettitekstien luotettavuutta ja hyödyntää perusteluita kirjoitelmassaan. Interventiossa oppilaille opetettiin tutkivan nettilukemisen taitoja (tiedonhaku, luotettavuuden arviointi ja synteesin laatiminen), minkä jälkeen taitoja harjoiteltiin kahdessa projektissa. Tulokset osoittivat, että nettitekstien luotettavuuden perusteleminen oli vaikeaa suurimmalle osalle kuudesluokkalaisista. Interventio lisäsi heidän taitojaan huomioida lähteiden piirteitä (esim. kirjoittaja ja julkaisupaikka) perusteluissaan.

Osatutkimuksessa II tutkittiin, miten lukiolaiset (N = 372) osasivat perustella nettitekstien luotettavuutta ja miten heidän Internet-spesifit episteemiset uskomuksensa olivat yhteydessä perustelutaitoihin. Tutkimus paljasti merkittäviä eroja lukiolaisten taidoissa käyttää eri luotettavuuden arviointikriteerejä (kirjoittaja, julkaisija, motiivit, evidenssi, korroboraatio) sekä arviointien syvällisyydessä. Lukiolaiset, jotka valitsivat hyödyllisempiä nettitekstejä sekä uskoivat arvioivansa kirjoittajaa ja vertaavansa eri tekstejä lukiessaan nettitekstejä, olivat parempia perustelemaan nettitekstien luotettavuutta.

Osatutkimuksessa III tutkittiin, miten tutkivan nettilukemisen interventio vaikutti lukiolaisten (N = 365) taitoihin tunnistaa, arvioida ja hyödyntää lähteiden piirteitä tutkivan nettilukemisen eri vaiheissa. Lukiolaiset tutkivat pienryhmissä valitsemaansa aihetta yhteisen työskentelydokumentin kysymysten ja tehtävien avulla. Opiskelijoiden työskentelyä tuettiin opettajan tietoiskuilla. Interventio lisäsi lähteiden piirteiden huomioimista, arvioimista ja hyödyntämistä hakukyselyissä, luotettavuusarvioinneissa sekä kirjoitelmassa. Interventiosta hyötyivät eniten ne opiskelijat, joiden kriittisen lukemisen taidot olivat heikoimmat ennen interventiota.

Avainsanat: tutkiva nettilukeminen, kriittinen lukeminen, luotettavuuden arviointi, Internet, interventio, peruskoulu, lukio

Author

Elina Hämäläinen
Department of Education
P.O. Box 35
FI-40014 University of Jyväskylä, Finland
elina.k.hamalainen@jyu.fi
Orcid: 0000-0001-7561-0530

Supervisors

Professor Miika Marttunen
Department of Education
University of Jyväskylä, Finland

Academic Research Fellow Carita Kiili
Faculty of Education and Culture
Tampere University, Finland

Senior Lecturer Eija Räikkönen
Faculty of Education and Psychology
University of Jyväskylä, Finland

Professor Paavo Leppänen
Department of Psychology
University of Jyväskylä, Finland

Reviewers

Professor Päivi Rasi-Heikkinen
Faculty of Education
University of Lapland, Finland

Professor Marc Stadtler
Institute of Educational Research
Ruhr University Bochum, Germany

Opponent

Professor Päivi Rasi-Heikkinen
Faculty of Education
University of Lapland, Finland

ACKNOWLEDGEMENTS

Over the twenty years that I have studied or worked at the University of Jyväskylä, I have had a kind of exciting and faraway dream that someday I will compose and defend my dissertation. Even though it took quite a long time to find the best moment to start the process, from that on, my Ph.D. journey has been quite fluent, precise, and efficient – as I am as a worker. But this dream could not have come true without other people actively helping, supporting, and encouraging me during these years.

First, I want to thank my main supervisor, Professor Miika Marttunen, for making it possible to start and go through this Ph.D. journey, trusting me as a researcher, and being there whenever I needed help. I am deeply grateful to my second supervisor, Dr. Carita Kiili, who has put so much effort into designing interventions for iFuCo and Aroni projects, supervising my dissertation articles, and supporting my career as a researcher. I have learned enormously from you in the areas of critical online reading and scientific writing. Carita, it has been such a joy that you share an enthusiastic interest in the same research area with me – even though your expectations for the quality of my manuscripts have sometimes been hard to reach. I also highly appreciate my third supervisor Dr. Eija Räikkönen. Your practical and scientific expertise in statistical analyses has raised my dissertation articles to a higher level and helped me find interest in that area. Eija, thank you also for all your encouraging feedback and kind words during this journey. Furthermore, I want to thank my fourth supervisor, Professor Paavo Leppänen, who made it possible to be involved in the iFuCo project and quite independently organize intervention and data collection procedures as a young researcher.

Further, I would like to express my significant appreciation to the reviewers, Professor Päivi Rasi-Heikkinen and Professor Marc Stadtler, for their careful and insightful evaluation of my dissertation, as well as Professor Rasi-Heikkinen, for agreeing to serve as the opponent during my public defense.

I want to thank my co-authors, Assistant Professor Roberto González-Ibañez, for your valuable expertise in creating the Neurone platform for the iFuCo project and Dr. Minna Lakkala and Dr. Liisa Ilomäki and Professor Auli Toom for efficient cooperation in designing an intervention for the Aroni project. I also want to thank all researchers in the iFuCo and Aroni projects, especially those who have taken care of the data collection and, of course, all participating adolescents and teachers. Thanks to Michael Freeman for carefully checking the language in my dissertation articles.

I have enjoyed being a member of our group of doctoral students with Minna, Timo, and Jukka. It has been interesting to share thoughts and challenges with you – of our dissertation articles and outside them. I am thankful for your continuing support. During the two decades, I have also had particularly warm-hearted and skillful colleagues at the University of Jyväskylä – in the Open University, Faculty of Sport and Health Sciences, Faculty of Information Technology, Department of Psychology, Department of Education, and

Department of Teacher Education. I appreciate your kind and encouraging words during my Ph.D. journey. I express my gratitude to the Academy of Finland and the Department of Education for funding my dissertation.

During the last few years, I have been honored to work on a Critical project funded by the Strategic Research Council. It has been a huge privilege to continue studying and enhancing the same research area as in my dissertation. I am also grateful to the project leaders, Kristian Kiili and Minna Torppa, for your support and encouragement to finalize my dissertation while working in Critical. It has been touching that so many researchers of the project have celebrated my achievements and been proud of my dissertation.

Finally, I thank my broader family for your support. My parents, Anja and Kari, have always highly appreciated education and, as teachers, directed my educational and research interest to pedagogical and school issues. During my whole life, my big sister Johanna has been a kind of role model for me, but this time, I was the first one of us to finalize a dissertation project. My dear friends and husband, Jyrki, thank you for your love, support, and laughter. You have helped me find a balance between research and other important – also relaxing – aspects of life. Despite this finalized dissertation to be proud of, my lovely children, Veera and Eetu, you are – and will be – my most significant achievements in life.

Jyväskylä 26.5.2023
Elina Hämäläinen

LIST OF PUBLICATIONS

This dissertation is based on the following publications, which are referred to as Sub-studies I, II and III in the text.

Article I 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

Article II 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

Article III 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

The author of this dissertation is the first author of all three research articles. She was involved in designing the interventions and their instructional materials at both educational levels. She was responsible for introducing half of the intervention group teachers at the sixth-grade level to the intervention lessons and materials, observing their lessons, and organizing the data collection in their classrooms. She participated in planning online inquiry tasks for both school levels and prior topic knowledge measure for upper secondary school students. She was responsible for analyzing the qualitative data, searching, and reviewing the literature, and writing the manuscripts. She conducted the statistical analyses with consultation from statistical expert (included as supervisor). The co-authors had advisory roles in the design of the studies and interpretation of the results, and they provided comments for all three manuscripts. Their role as co-authors also included writing editing and reviewing comments for manuscripts. The data used in the three publications has been collected as part of two research projects funded by the Academy of Finland, iFuCo project and Aroni project.

TABLES

CONTENTS

1 INTRODUCTION

During the last few years, our feelings of security have been disrupted owing to the Covid-19 pandemic and the Russian attack on Ukraine. These events have also remarkably affected the online information environment around us (e.g., Pavlik et al., 2022; Roozenbeek et al., 2020). As we desperately search for information on the Internet to understand what is happening, the rapid growth of misinformation has challenged our ability to distinguish between more credible and less credible information. Some stakeholders have motives to intentionally share disinformation (e.g., Lewandovsky et al., 2013), but more often, it might be a question of information about complex phenomena that must be carefully explored (cf. Ecker et al., 2022). In the new circumstances, even experts may struggle to discover what information can be trusted (Stewart, 2021).

The spread of misinformation and disinformation on the Internet is a remarkable societal problem as it can lead, for example, to growing inequality and polarization among people (cf. Lewandowsky et al., 2017). In Finland, most citizens are able to access the Internet, although the oldest people are most often Internet nonusers (Rasi, 2018). However, not all citizens possess abilities to access, read, and understand credible information, and thus, some are particularly vulnerable to misleading online information (cf. Paakkari et al., 2022). For example, a person's lower socioeconomic status has been associated with the belief in conspiracy theories (see review by Tsamakis et al., 2022) and weaker online reading skills (Leu et al., 2015). Because the spread of false information on the Internet is quite difficult to affect, the basic skills needed to search, evaluate, and interpret (online) information should be taught already during basic education, which reaches all children from different backgrounds.

The critical evaluation of information has been connected to three of eight competencies for lifelong learning proposed by the Council of the European Union (2018). Searching for reliable information is more difficult nowadays, as the Internet allows everyone to publish their claims and ideas. Accordingly, a tremendous amount of information of varying quality is available (e.g., Salmerón et al., 2018b). Online information can also be rapidly and easily shared without

expressing its origin or source. Thus, familiarizing oneself only with the content can be misleading and can predispose a reader to non-credible information (McGrew & Byrne, 2020). In all, uncritical reading may lead to trust in misinformation when making important decisions regarding, for example, health, environment, or societal issues (cf. Ecker et al., 2022).

The newest curricula in Finland emphasize learning of critical thinking skills and skills to evaluate information (National Core Curriculum for Basic Education, 2014; National Core Curriculum for General Upper Secondary Education, 2019). Although most adolescents read information on the Internet for entertainment purposes (e.g., Macedo-Rouet et al., 2020), online texts are increasingly being used for school tasks and assignments. A traditional textbook is accessible to teachers and students because it often coherently defines the arguments and findings. Instead, when students independently search for information on the Internet, they must spontaneously evaluate and compare information in multiple online texts (e.g., Macedo-Rouet et al., 2013). Compared with using traditional textbooks, this information-seeking method is more demanding for students, who often select the most easily found documents (cf. Haas & Unkel, 2017; Pan et al., 2007). However, the process of online inquiry (cf. Leu et al., 2019) might be as important as its result when during the process, students engage in sourcing and corroborating information on multiple documents (cf. McGrew & Byrne, 2022). Thus, they can more deeply learn about the differences between more and less credible online texts.

Furthermore, when students engage in online inquiry, it challenges teachers' abilities to guide their process to find and recognize credible information. Students need to know, for example, how to search for information online, evaluate its credibility, and synthesize information from multiple online documents (cf. Leu et al., 2019). Unfortunately, research has shown that many early adolescents (e.g., Forzani, 2018; Kanniainen et al., 2022) and secondary school students (Barzilai et al., 2015; McGrew et al., 2018) lack these skills, although they have been considered as diginatives. During recent decades, promising interventions have been developed to teach these skills to students at different educational levels (see reviews by Brand-Gruwel & van Strien, 2018; Brante & Strømsø, 2018; Bråten et al., 2018c). Hence, efficient teaching methods should not only cover the explicit teaching of online inquiry skills but also allow students' thinking, discussions, and collaboration with peers. Further, the tasks and investigated topics should be not only interesting for students but also sufficiently problematic to require critical reading abilities.

In this dissertation, critical online reading refers to considering, evaluating, and synthesizing source information and content of online texts during online inquiry. Although various interventions to enhance students' critical online reading skills have been conducted during the last decades (see reviews by Brand-Gruwel & van Strien, 2018 and Brante & Strømsø, 2018), only few of these interventions have covered the entire process of online inquiry (cf. Argelagós & Pifarré, 2012; Kingsley et al., 2015). Further, most interventions have been conducted among university students, and less is known about how to teach

critical online reading skills to younger students. This dissertation aims to develop efficient instructional methods for adolescents and test the efficacy of those methods. To further develop teaching methods, more detailed information is needed about adolescents' critical online reading skills and learning of those skills with related individual differences and other factors.

Therefore, this dissertation investigated whether interventions implemented by teachers in primary and upper secondary school can enhance adolescents' critical online reading skills, particularly their abilities to justify the credibility of online texts and engage in sourcing. Further, this study provided information about adolescents' critical online reading skills and learning of those skills during the interventions. Thus, the dissertation aims to further advance reading research from the perspective of educational psychology. However, there are also joint viewpoints with information and communication sciences, media education, and youth research, for example.

2 THEORETICAL AND CONCEPTUAL FRAMEWORKS

This chapter describes the theoretical and conceptual frameworks for critical online reading. Critical online reading can be based on theoretical models describing the phases of online inquiry (Leu et al., 2019), multiple document reading (Perfetti et al., 1999), and reciprocal evaluation strategies (Barzilai et al., 2020). Conceptually, critical reading shares history and similarities, for example, with critical literacy and critical thinking, although it also differs from them (Cervetti et al., 2001). Furthermore, critical online reading skills are also represented in other literacy concepts, such as information literacy (Zurkowski, 1974), media literacy (see Potter, 2022), and multiliteracy (New London group, 1996).

2.1 Theoretical frameworks

This dissertation relies on three theoretical frameworks: the online research and comprehension model (Leu et al., 2019), the documents model (Perfetti et al., 1999), and the bidirectional model of first- and second-hand evaluation strategies (Barzilai et al., 2020). First, the online research and comprehension model by Leu et al. (2019) describes the phases of online inquiry that a critical reader encounters when investigating a topic on the Internet. Second, the documents model (Perfetti et al., 1999; Rouet, 2006) demonstrates how critical readers acknowledge, evaluate, and compare source information and content in multiple texts to understand the examined topic. Third, the bidirectional model by Barzilai et al. (2020) presents a repertoire of evaluation strategies that readers can use to reciprocally judge the credibility of source information and content. Thus, the first model builds the structures and phases for online inquiry, and the other two models accentuate the role of sourcing and evaluation during reading. In the following sections, the importance of these models for this dissertation and the relations between the models are presented in more detail.

2.1.1 The online research and comprehension model

Online inquiry refers to a problem-based process where the reader is goal-oriented when investigating a topic on the Internet (cf. Leu et al., 2019). According to the online research and comprehension model (Leu et al., 2019), the process of online inquiry comprises five key phases: formulating questions and searching for, evaluating, synthesizing, and communicating information. Furthermore, these phases can be seen as reciprocal—affecting each other—and the entire process forms a continuing cycle. Thus, critical online reading, including the evaluation processes (see Gerjets et al., 2011) and sourcing practices (see Bråten et al., 2018c), should play a role in every phase of online inquiry (see also Kiili et al., 2021).

When online readers formulate questions for their inquiry or specify their information need (Leu et al., 2019), they should carefully consider and evaluate what information they need to solve the problem. During this phase of online inquiry, readers can underline the use of credible information (cf. Gerjets et al., 2011) and even specify reliable authors or organizations (source information) that may write about the topic under investigation (see Kiili et al., 2021). In the following search phase, they can apply these specific considerations when formulating queries for search engines (Leu et al., 2019). In addition to relevant content-related search terms, readers can use, for example, the names of credible authors and organizations (source information) in their search queries (cf. Kiili et al., 2021).

Evaluation can continue when online readers receive the search results page. When selecting which online texts are worth reading more closely, they can evaluate relevance and credibility based uniform resource locator (URL) addresses, titles, and example texts (e.g., Hahnel et al., 2020; Rieh, 2002). After opening a particular link on the search results page, readers can more specifically assess the source information and content of the online text by paying attention to, for example, the author, venue, and purpose of the text (cf. Perfetti et al., 1999; Rouet, 2006) as well as the claims and evidence presented in the text (e.g., Forzani, 2020; Sinatra & Lombardi, 2020). Thus, the evaluation focuses on the source information and content of the texts as well as on how these are related to each other (cf. Barzilai et al., 2020; Stadtler & Bromme, 2014).

In the synthesizing phase, the critical reader compares selected online texts by investigating their similarities and differences in relation to the topic under investigation (Leu et al., 2019). However, even in this phase, evaluating and contrasting source information and content as well as the different sources of the texts is crucial (cf. Gerjets et al., 2011). Accordingly, skillful readers formulate an intertext model including source-content and source-source links (see Perfetti et al., 1999). In the final stage of online inquiry, the findings are often communicated to others (Leu et al., 2019) through, for example, a written product or oral presentation.

2.1.2 The documents model

The documents model (e.g., Britt et al., 2018; Perfetti et al., 1999) describes how skillful readers can synthesize information from multiple documents. It also accentuates the role of sourcing in multiple document reading. When readers build the documents model, the processes include investigating how statements or the entire content of different documents relate to each other (intertextual connections), presenting source information of the documents (e.g., publication venue or name of the author), and expressing which statement or content is stemmed from which source. Thus, the model (Perfetti et al., 1999; Rouet, 2006) suggests that skillful readers can build two types of mental representations: an intertext and an integrated model.

When building an intertext model, readers can construct source-content and source-source links (Perfetti et al., 1999; Rouet, 2006). Source-content links contain the source information of a document (e.g., author, publisher, and intentions) and its content. In contrast, by formulating source-source links, readers can connect sources (e.g., authors) from multiple documents by showing their relationships, which can be, for example, supporting or opposing by nature. The intertext model is essential for situations where readers encounter texts including conflicting information that they cannot coherently integrate. Thus, they need to discern which texts can be trusted (Britt et al., 2014) and which texts contradict the reliable ones. On the Internet, at least partially disagreeing information often exists; thus, the intertext model needs to be constructed.

Furthermore, when readers combine the contents of multiple documents to understand the topic under investigation, they create an integrated mental model (Perfetti et al., 1999; Rouet, 2006). The wide range of conflicting online information has made it difficult to coherently combine it (e.g., Saux et al., 2021). However, readers can form a coherent mental representation without excluding contradicting information by taking advantage of the intertext model (e.g., Rouet et al., 2016; Saux et al., 2021). When readers interconnect their intertext and integrated mental models by informing who said what and by using this information to evaluate and interpret the content of each document (Britt et al., 2014), they discover the entire documents model (Perfetti et al., 1999).

2.1.3 The bidirectional model of first- and second-hand evaluation strategies

The bidirectional model of first- and second-hand evaluation strategies by Barzilai et al. (2020) (see also Stadtler & Bromme, 2014) proposes that people can use versatile first- and second-hand evaluation strategies to judge the credibility of information. When facing contradicting information, most experts rely on first-hand evaluation strategies by attempting to evaluate information validity (Stadtler & Bromme, 2014). However, first-hand strategies can be challenging for laypersons when they lack specialized topic knowledge and abilities to evaluate presented claims and evidence (e.g., Bromme & Goldman, 2014). Therefore, second-hand evaluation strategies referring to sourcing offer ways to evaluate

more accessible source features such as author expertise and intentions (see Bromme & Goldman, 2014).

According to this model (Barzilai et al., 2020), first-hand evaluation strategies include knowledge-based validation, discourse-based evaluation strategies, and corroboration. Knowledge-based validation occurs when readers judge the information in light of their prior knowledge and beliefs about the topic. For example, previous research has shown that people tend to mainly rely on information consistent with their own beliefs (e.g., Murphy & Alexander, 2004). Discourse-based evaluation strategies are based on various discourse features, such as the consistency of arguments presented in the text (e.g., von der Mühlen et al., 2016) or the writing style (e.g., Bromme et al., 2015). In particular, considering evidence can refer to evaluating the quality and balance of argumentation (e.g., Iordanou et al., 2019) as well as evaluating what kind of evidence (e.g., research, own experiences) is presented to support the claims in the text (e.g., Hoeken, 2001). Finally, corroborating the content of information with other documents is essential when judging its credibility (Barzilai et al., 2020). Corroboration involves comparing the information from various documents to identify which statements are agreed upon and which are discrepant (Britt & Aglinskas, 2002).

In the bidirectional model (Barzilai et al., 2020), second-hand evaluation strategies refer to sourcing. According to Bråten et al. (2018c), sourcing refers to attending to, representing, evaluating, and using information about the sources of textual content. Sourcing is particularly important when documents include conflicting information about the topic under investigation, and readers must decide whom and what to trust (e.g., Stadtler & Bromme, 2014). Thus, it helps, for example, to attend to reliable sources, evaluate the credibility of information, and use source information from multiple documents to coherently understand the topic (Bråten et al., 2018b). In addition, sourcing saves time from reading untrustworthy information when the author or publisher is initially evaluated as highly unreliable (McGrew & Byrne, 2020). However, evaluating source information (e.g., author expertise, intentions) requires that it is available to and understandable for readers. Furthermore, readers often neglect source information while reading or fail to accurately interpret it (e.g., Macedo-Rouet et al., 2019).

Finally, the model by Barzilai et al. (2020) highlights the reciprocal relationship between judgments about source features and content. It suggests that source evaluation strategies (sourcing) inform source trustworthiness judgments and can indirectly influence judgments about the validity and quality of content. Conversely, content evaluation strategies inform content validity and quality judgments and can indirectly influence judgments about source trustworthiness.

In this dissertation, "critical online reading skills" refer to readers' abilities to consider and evaluate the trustworthiness of the sources and validity of the text contents (Barzilai et al., 2020) when engaging in different phases of online

inquiry (Leu et al., 2019) to form a coherent representation of the topic under investigation (Perfetti et al., 1999).

2.2 Conceptual framework

As the definition of critical online reading skills at the end of section 2.1.3 indicates, in this dissertation, critical reading is contextualized on the Internet. Through the decades, critical reading has been considered to include essential abilities such as considering the authors' purpose, distinguishing opinions from facts, making inferences, and forming judgments (see Cervetti et al., 2001). It has also been linked to critical literacy, emphasizing that texts are understood in the context of social, historic, and power relations and that critical reading aims to find means for social transformation (e.g., Cervetti et al., 2001; see also Freire, 1985; New London Group, 1996). When approached from the liberal-humanist tradition (Cervetti et al., 2001), critical reading focuses on understanding authors' intentions and interpreting whether the information is valid or worthy of skepticism. In critical reading, the processes of sense-making, deduction, or rational analysis have been emphasized. In addition, it has been seen that facts, inferences, and reader judgments can be separated from each other during critical reading (Cervetti et al., 2001).

For critical reading, the information on the Internet poses more challenges than traditional texts. Because the Internet does not include traditional gatekeepers, online readers are responsible for determining the origin, motives, and sources of information (Salmerón et al., 2018b). In this dissertation, "critical online reading" is defined as considering, evaluating, and synthesizing the source information and content of multiple online texts (cf. Perfetti et al., 1999) during different phases of online inquiry (cf. Leu et al., 2019). Further, the reciprocal relationship between the credibility evaluation of content and source information in online texts is acknowledged (cf. Barzilai et al., 2020; Stadtler & Bromme, 2014). For coherence, the term "online text" is used throughout this dissertation to refer to researcher-designed texts or authentic online texts. Furthermore, the concept of the text refers to static multimodal texts, including written language and visuals, compared to more versatile symbol systems and their combinations related to the multiliteracy concept in the Finnish education system curricula (see Rasi et al., 2019).

In this dissertation, sourcing is considered as a part of the practices of students specifying their information need, formulating search queries, and composing a written product (see also Kiili et al., 2021). Similarly, other evaluative practices can occur in different phases of online inquiry (e.g., Gerjets et al., 2011; see also Kiili et al., 2021). Furthermore, evaluative and sourcing practices are considered iterative and reciprocal so that they are intertwined in different phases of online inquiry (cf. Abed & Barzilai, 2023; Forzani et al., 2022; Kiili et al., 2021). In this dissertation, adolescents' critical online reading skills are

measured as individual skills but are also supported through collaborative reading practices.

Aside from measuring students' critical online reading skills, this dissertation measured older adolescents' epistemic beliefs regarding how students believe that they evaluate the credibility of online texts. Epistemic beliefs, particularly the justifications for knowing, reflect the ways readers suppose they, for example, evaluate the credibility of knowledge claims and decide whom and what to believe (Sandoval et al., 2014). Greene et al. (2008) proposed that justifications for knowing cannot be displayed by a single dimension. Accordingly, they suggested two dimensions: "justification by authority" and "personal justification." A third dimension, "justification by multiple sources," was found in the think-aloud study by Ferguson et al. (2012). These three knowing dimensions have been incorporated into an Internet-specific inventory for students' epistemic beliefs (Bråten et al., 2019a), which has been applied in this dissertation to measure how older adolescents' epistemic justification beliefs were associated with their critical online reading skills.

Furthermore, critical reading relates to critical thinking, when clear and logical analysis is emphasized (Cervetti et al., 2001). In the Finnish National Core Curriculum for Basic Education (2014), critical thinking, critical consideration, or critical evaluation is mentioned in the aims of almost all school subjects and related to many transversal competencies (e.g., multiliteracy); it is also referenced in the value basis of the entire curriculum. These aims are further highlighted at the upper secondary school level (National Core Curriculum for General Upper Secondary Education, 2019). In addition, source criticism and evaluation of knowledge bases are emphasized more at the upper secondary school level than in basic education. Moreover, in the descriptions and aims of language arts, aspects of critical reading are most apparent (see National Core Curriculum for Basic Education, 2014; National Core Curriculum for General Upper Secondary Education, 2019), even though the curricula do not use the term "critical reading". Interestingly, the Internet or "online information" is rarely mentioned in the curricula.

Critical (online) reading can also be approached from overlapping literacy frameworks, such as information literacy, media literacy, and multiliteracy, even though there is a lack of shared definitions of these multifaceted concepts (cf. meta-review of Wuyckens et al., 2022). The term "information literacy" derives from informational sciences (Zurkowski, 1974). After the emergence of online information, Kuhlthau (1991) developed and validated the Information Search Process (ISP) model comprising similar strategies during a cyclic process as the online research and comprehension model by Leu et al. (2019). However, among information sciences, locating information has been more highlighted than the later phases of online inquiry.

The term "media literacy" includes, among other skills, exposure skills (e.g., searching, selecting) and information processing skills (e.g., critical reading, evaluating, synthesizing) (see review by Potter, 2022) that are closely related to critical online reading skills. Note that the concept of media is much broader than

"online text" utilized in this dissertation. As mentioned above, Finnish national core curricula for schools apply the term "multiliteracy" (cf. New London group, 1996), defined briefly as communication abilities, such as interpreting, producing, and making a value judgment across a range of different texts (see Rasi et al., 2019). However, it has been concluded that the concept used in research differs from the multiliteracy concept applied in the Finnish core curricula (see Palsa & Ruokamo, 2015). Nevertheless, the latter includes skills to search, evaluate, and interpret information alongside critical thinking (see National Core Curriculum for Basic Education, 2014; National Core Curriculum for General Upper Secondary Education, 2019), similarly with the critical online reading skills.

3 PREVIOUS RESEARCH

Based on previous studies, this chapter describes what is known about students' critical online reading skills and how individual differences are connected to their skills. Further, interventions targeted to improving students' critical online reading skills are reviewed.

3.1 Students' critical online reading skills

Critical online reading skills have been investigated among students in different age groups: young adolescents (e.g., Coiro et al., 2015; Forzani, 2018; Forzani et al., 2022; Kanniainen et al., 2019, 2022; Kiili et al., 2018b, 2023), lower secondary school students (e.g., Abed & Barzilai, 2023; Macedo-Rouet et al., 2020; Walraven et al., 2009), upper secondary school students (e.g., Kiili et al., 2008, 2019, 2022a; Marttunen et al., 2021; McGrew et al., 2018), and university students (e.g., Barzilai et al., 2015; Hahnel et al., 2020; Tsai et al., 2022). This section describes the results of the studies focused on investigating young adolescents' and upper secondary school students' critical online reading skills.

For young adolescents, critical online reading is particularly challenging. For example, the qualitative analysis by Coiro et al. (2015) revealed that most seventh graders recognized the author of an online text but evaluated the author expertise in irrelevant, vague, and superficial ways. Furthermore, young adolescents had difficulties justifying the overall credibility of online information in a reasoned manner. Forzani (2018) found that during the process of online inquiry, most seventh graders (N = 1434, in total) did not perform well in locating, evaluating, synthesizing, and communicating information. Evaluating online information was particularly difficult for students. In line with these studies, Kiili et al. (2018b) found that many Finnish sixth graders (N = 426) had limited abilities to justify the credibility of online texts. However, note that the variation between students' skills was high.

The variation in students' critical online reading skills has been more evident among older students. Accordingly, the think-aloud study by Kiili et al. (2008), exploring Finnish upper secondary school students' evaluation skills, revealed the wide variation between students' skills. The most versatile evaluators used various evaluation strategies and mainly focused on credible online texts. However, uncritical readers spent much time reading less credible online texts and could not constructively use evaluation strategies. A decade later, McGrew et al. (2018) conducted a study among middle school, high school, and college students (N = 894). They found that students did not often evaluate the author of the online texts, made judgments about the credibility based on surface features, were satisfied with shallow information, and failed to evaluate evidence found in online texts. In a recent study by Kiili et al. (2022a), Finnish upper secondary school students' (N = 73) credibility evaluations were entirely accurate for author, intentions, venue, and evidence of online texts, but their credibility justifications lacked sophistication. Following previous studies, inter-individual differences were also considerable.

Despite being an essential component of critical online reading skills, sourcing seems to be rarely a spontaneously used skill and challenging for students of different ages (Bråten et al., 2018c). As shown above, young adolescents can struggle with identifying the author or publisher of the text (Coiro et al., 2015; Kiili et al., 2018b), whereas many older students have difficulties in recognizing or evaluating the intentions of the author or publisher (e.g., Potocki et al., 2020) as well as citing and comparing sources of the texts when composing a written product (e.g., Kiili et al., 2020; Pérez et al., 2018). For example, Paul et al. (2017) found that students tend to overestimate their sourcing skills. On the other hand, students may possess skills to evaluate source information, but they do not apply these skills, for different reasons, in reading situations (Paul et al., 2017).

To conclude, considerable differences in students' critical online reading skills exist through different school levels, even though these skills are more challenging for younger than for older students. However, older students lack the skills to justify the credibility of online texts and regularly engage in sourcing. Therefore, investigating the factors that might explain differences between students' critical online reading skills and developing efficient methods to teach these skills is important.

3.2 Individual differences related to students' critical online reading skills

Several factors, such as cognitive and affective, can cause individual differences in reading and reading comprehension (Afflerbach, 2016) and, respectively, be associated with students' critical online reading skills (cf. Barzilai & Strømsø, 2018). Thus, they may also affect the learning of those skills. In the recent review

(Anmarkrud et al., 2021), individual differences in the conceptualizations of multiple document representation and use were categorized into reading skills and strategies, cognitive factors, motivation and engagement, beliefs, personality, and expertise. Among these individual difference factors, this dissertation focuses on students' basic reading skills, prior topic knowledge, and epistemic beliefs, which are an essential part of the conceptualizations and are highly acknowledged in previous studies (see Anmarkrud et al., 2021). Furthermore, of the sociodemographic factors (e.g., parental educational level and socioeconomic status) gender differences at the primary school level are explored.

Because students' basic reading skills form the basis for their critical online reading skills (e.g., Kanniainen et al., 2019, 2022), various studies have investigated the association between these skills (Anmarkrud et al., 2021). In the literature, students' reading fluency has been associated with young adolescents' abilities to evaluate source credentials (Macedo-Rouet et al., 2013) and with high school students' abilities to discriminate between reliable and unreliable links (Macedo-Rouet et al., 2020). Forzani (2018) and Kiili et al. (2018b) also found that young adolescents with better basic reading skills were better evaluators than others. Further, studies measuring students' reading comprehension using open-ended questions have often found a positive relationship between reading comprehension and critical reading skills (e.g., Hahnel et al., 2019; Macedo-Rouet et al., 2013; Salmerón et al., 2020). That is, open-ended questions are often more difficult for students to answer than, for example, cloze tests (see Kullberg et al., 2022) or multiple-choice tests and require the construction of mental representations that align with building a document model (cf. Perfetti et al., 1999).

Regarding individual differences, previous studies have most often examined students' prior topic knowledge in relation to their critical reading skills (Anmarkrud et al., 2021). However, although various studies have measured the association, no consistent findings have been achieved. For example, studies measuring students' prior topic knowledge with true/false items have not usually found relationship with their critical reading skills (e.g., Kammerer et al., 2016b; Ulyshen et al., 2015). Similarly, Kiili et al. (2022a) found that upper secondary school students' prior topic knowledge was not associated with their credibility evaluation skills. The review by Anmarkrud et al. (2021) revealed that students' prior topic knowledge more often had a relationship with their sourcing skills when students investigated a science topic compared with when they explored a health topic.

Students may also differ in how their beliefs about evaluating information when reading it relate to their evaluations. Epistemic beliefs such as justifications for knowing (e.g., verifying knowledge claims by assessing the author or by comparing multiple documents) and their relationship with critical reading skills have been mainly studied among university students (Anmarkrud et al., 2021). Justifications for knowing have shown correlations with students' author evaluations and predicted their trustworthiness ratings of the texts (Strømsø et

al., 2011). Further, in an eye-tracking study (Kammerer et al., 2013), students' Internet-specific justifications for knowing negatively correlated with and predicted the number of utterances concerning parts of the search engine results pages (SERPs). In particular, justification by authority was positively correlated with and indicated students' source evaluation comments during a think-aloud study in the online environment (Kammerer et al., 2021). However, in the study by Kiili et al. (2022a), upper secondary school students' Internet-specific epistemic justifications (ISEJ) were not associated with their credibility evaluation skills.

Although girls traditionally have better basic reading skills than boys, particularly in Finland (e.g., Programme for International Student Assessment (PISA) studies by Harju-Luukkainen et al., 2016; Leino et al., 2018; see also Progress in International Reading Literacy Study (PIRLS) 2016 by Marôco, 2021), results regarding differences in their critical online reading skills have been mixed. Forzani (2018) and Kiili et al. (2018b) found that among young adolescents, girls outperformed boys in evaluating online information. However, Kanniainen et al. (2019) found mixed results, as gender did not explain the differences when young adolescents confirmed the credibility of online texts, but girls were better than boys in questioning the credibility of online texts.

As most findings regarding the role of individual differences were mixed, Anmarkrud et al. (2021) suggested in their review that the associations between students' individual differences and critical reading skills seem to depend on how they are measured and which topic or domain the reading materials address. Interestingly, studies have found associations between individual differences and students' sourcing skills more often when students have been prompted to source (e.g., responded to questions) than during spontaneous sourcing, such as using citations in their essays (Anmarkrud et al., 2021).

3.3 Previous interventions targeting students' critical online reading skills

As students with individual differences may struggle with their critical online reading skills, a growing number of interventions for improving their skills have been conducted at different educational levels (see reviews by Brand-Gruwel & van Strien, 2018; Brante & Strømsø, 2018; Bråten et al., 2018c). However, these interventions have considerably varied in length, content, and measured outcomes. Furthermore, few interventions have been conducted at the primary school level (e.g., Kingsley et al., 2015; Macedo-Rouet et al., 2013; Zhang & Duke, 2011), and only a few interventions (e.g., Argelagós & Pifarré, 2012; Kingsley et al., 2015) have covered the entire process of online inquiry. Overall, the interventions among upper secondary school students have achieved primarily positive results (e.g., Braasch et al., 2013; Bråten et al., 2019b; Britt & Aglinskas, 2002; McGrew, 2020; McGrew & Byrne, 2020), whereas results have been mixed

among primary school students (e.g., Kingsley et al., 2015; Macedo-Rouet et al., 2013; Zhang & Duke, 2011).

Despite the differences mentioned above, the interventions have applied similar instructional methods. As many students seem to be aware of the need for critical reading skills but unsure of how to use them (Paul et al., 2017), it seems evident that strategies need to be explicitly taught (e.g., Heijltjes et al., 2014; Marin & Helpern, 2011) or modeled (e.g., Coiro, 2011b) for them. Students' spontaneous sourcing has also been quite rare among adolescents (e.g., Walraven et al., 2009); therefore, prompts (e.g., Kammerer et al., 2016b; Macedo-Rouet et al., 2019), including questions and sub-tasks in students' worksheets or digital working environments, can be used to motivate them to regularly practice the skills. Previous studies have also shown that the contrasting cases approach (Braasch et al., 2013) and the use of multiple documents (e.g., Bråten et al., 2019b) can be beneficial for learning and practicing critical reading skills. Along with modeling and practice, discussions with peers and a teacher are essential for students to express their thinking and learn from others' ideas (see Bråten et al., 2019b; Macedo-Rouet et al., 2013; Péréz et al., 2018). Similarly, students can benefit from collaborative work (e.g., Kiili et al., 2019), but individually practicing skills is also important (cf. Frerejean et al., 2018).

In the interventions of this dissertation, the above-introduced instructional methods have been developed and combined to scaffold students through the entire process of online inquiry.

4 RESEARCH QUESTIONS

This dissertation aims to develop methods to analyze adolescents' critical online reading skills and clarify the role of individual differences and topic-related factors on their skills. Moreover, it was aimed to develop instructional methods and materials for enhancing adolescents' critical online reading skills and test the efficacy of those methods. Although students' skills were examined and taught in school contexts, students can apply similar skills in their leisure time when consuming online information. This dissertation comprises three sub-studies. Sub-study I examined whether intervention promoted sixth graders' justifications for the credibility of online texts. Sub-study II investigated upper secondary school students' abilities to justify the credibility of online texts and how students' epistemic beliefs were associated with their evaluation skills. Further, the purpose of Sub-study III was to clarify how the intervention affected upper secondary school students' sourcing skills. This dissertation addressed the following research questions.

RQ1: What kinds of critical online reading skills did students have?

a) How did sixth graders (Sub-study I) and upper secondary school students (Sub-study II) justify the credibility of online texts?
b) How were sixth graders' justifications for the credibility of online texts reflected in their written product? (Sub-study I)
c) How did upper secondary school students engage in sourcing during the online inquiry task? (Sub-study III)

RQ2: Did the intervention enhance students' critical online reading skills?

a) Did the intervention enhance sixth graders' justifications for the credibility of online texts and the use of those justifications in the written product? (Sub-study I)
b) Did the intervention increase upper secondary school students' sourcing during online inquiry? (Sub-study III)
c) How did upper secondary school students' sourcing skills change during the intervention? (Sub-study III)

RQ3: How were individual difference factors such as students' gender (Sub-study I), basic reading skills (Sub-studies I–III), prior topic knowledge (Sub-studies II and III), and Internet-specific epistemic justifications (Sub-study II) associated with their critical online reading skills?

RQ4: Was the topic (Sub-study II) or topic-order (Sub-studies I and III) of the online inquiry task and students' text selections (Sub-study II) associated with their critical online reading skills?

5 METHODS

Data for this dissertation were collected in two research projects funded by the Academy of Finland: (1) Enhancing learning and teaching for future competences of online inquiry (iFuCo 2016–2018; number 294197) and (2) Argumentative online inquiry in building students' knowledge work competence (Aroni 2015–2019; number 285817). Projects were interdisciplinary, including researchers from educational sciences, information science, and psychology. An intervention study was conducted among sixth graders (iFuCo) and upper secondary school students (Aroni). Table 1 summarizes the research questions, research designs, participants, data, and data analysis of the three sub-studies. In the following sections, the research designs and data employed in the sub-studies are presented in more detail.

TABLE 1 Summary of the Research Questions, Research Design, Participants, Data, and Data-Analysis of Sub-studies I–III

	Sub-study I: Promoting sixth graders' credibility evaluation of Web pages: An intervention study	**Sub-study II: Students' abilities to evaluate the credibility of online texts: The role of Internet-specific epistemic justifications**	**Sub-study III: Teaching sourcing during online inquiry: Adolescents with the weakest skills benefited the most**
Research questions	1. How did the sixth graders evaluate the credibility of Web pages? 2. Did the teacher-led intervention lead to improvement in the sixth graders' performance on an online credibility evaluation task compared to the control group? 3. How were students' credibility evaluations reflected in their written products? Did the teacher-led intervention result in increase of students' use of justifications for credibility in their written products?	1. How well did students evaluate the credibility of self-selected online texts when provided with a range of online texts via Google Custom Search Engine? 2. How were students' Internet-specific epistemic justifications associated with their evaluation performance when the usefulness of text selections, reading fluency, and prior topic knowledge were controlled for? 3. Did the associations between students' Internet-specific epistemic justifications and their evaluation performance differ according to the topic?	1. Did upper secondary school students' sourcing in different phases of an online inquiry through a teacher-led intervention increase compared to controls? 2. How did students' sourcing performance change during the intervention? 3. 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?
Research design	Quasi-experimental pre-post design with a non-equivalent control group	Cross-sectional and explanatory	Quasi-experimental pre-post design with a non-equivalent control group
Participants	Finnish sixth graders Intervention group (N = 192) Control group (N = 150)	Finnish upper secondary school students (N = 372)	Finnish upper secondary school students Intervention group (N = 196) Control group (N = 169)

continues

Table 1 continues

	Sub-study I: Promoting sixth graders' credibility evaluation of Web pages: An intervention study	Sub-study II: Students' abilities to evaluate the credibility of online texts: The role of Internet-specific epistemic justifications	Sub-study III: Teaching sourcing during online inquiry: Adolescents with the weakest skills benefited the most
Qualitative data	Students' responses to the online inquiry tasks: their written justifications for credibility ratings of Web pages and written products (synthesis)	Students' responses to the online inquiry task: their text selections and written justifications for the credibility of online texts	Students' responses to the online inquiry tasks: their sourcing in specifying information need, search queries, credibility judgments, and written product (synthesis)
Quantitative data	- Reading fluency and comprehension tests	- Internet-specific epistemic justifications (ISEJ) inventory - Reading fluency test - Prior topic knowledge test	- Reading fluency test - Prior topic knowledge test
Data analysis		- Scoring of students' responses (content analysis) - Confirmatory factor analysis (CFA) and structural equation modeling (SEM)	- Scoring of students' responses (content analysis) - Linear and negative binomial regression analyses and reliable change index (RCI)

5.1 Participants and research designs

In both, iFuCo and Aroni research projects, an intervention with a quasi-experimental pre-post design and a non-equivalent control group were conducted (cf. Handley et al., 2018). The school teachers implemented the interventions instead of the researchers, that is referred with a term "teacher-led" throughout this dissertation. The interventions were adjusted in the regular school curricula, resulting in all students completing tests and tasks. However, the responses of only those students who gave informed consent were used for research purposes. Guardian(s) also gave consent for underage students.

In Sub-study I, data from the intervention study among sixth graders (iFuCo) were employed. Altogether, 342 sixth graders (M = 12.13; SD = 0.41) participated in the Sub-study I. Further, cross-sectional data from the pre-measurement phase (Aroni) were utilized in Sub-study II. In total, 372 upper secondary school students (M = 17.35; SD = 0.40) participated in Sub-study II. Finally, data from the intervention study among upper secondary school students (Aroni) were applied in Sub-study III. Altogether, 365 upper secondary

school students (M = 17.35; SD = 0.40) participated in the Sub-study III, comprising the same students as in Sub-study II. According to the Official Statistics of Finland (2020, 2022), the gender distribution of the participants corresponds to that in Finnish basic education (Sub-study I; 48% females) and students graduating from upper secondary school in Finland (Sub-studies II and III; 59% females).

In the intervention studies conducted in Sub-studies I and III, students were divided into an intervention group and a control group. For practical reasons, the intervention group teachers (eight class teachers in Sub-study I and five language arts teachers in Sub-study III) were recruited based on their opportunity and willingness to implement intervention lessons. Control group teachers (seven class teachers and six language arts teachers) were recruited after the selection of the intervention group teachers. Thus, in Sub-study I, 192 sixth graders in eight classes formed the intervention group, and 150 sixth graders in seven classes formed the control group. Similarly, in Sub-study III, 196 upper secondary school students in nine courses formed the intervention group, and 169 upper secondary school students in seven courses formed the control group.

In the interventions, all students performed an online inquiry task as a pre- and post-test. For all students, the topic under investigation was different in the pre- and post-tests. Between the tests, the intervention group of sixth graders (Sub-study I) participated in the intervention program on online inquiry skills (21 × 45 min lessons within six weeks) during regular schoolwork. In contrast, the intervention group of upper secondary school students (Sub-study III) received an online inquiry intervention (4 × 75 min lessons within 1.5 weeks) as part of their Texts and Influence course. At the time of their course, most upper secondary school students had a second year of their studies. Respectively, the control group of sixth graders followed business-as-usual teaching. At the upper secondary school level, the control group participated in a regular Texts and Influence course. All control group teachers in both interventions received intervention materials after the study was completed.

5.2 Ethical considerations

All sub-studies of this dissertation were part of research projects funded by the Academy of Finland. Thus, they followed the ethical guidelines of the Finnish National Board on Research Integrity (2019) and those of the consortium universities, including the general ethical principles of respecting the dignity and autonomy of participants and material and immaterial cultural heritage and biodiversity and avoiding causing significant risks, damage, or harm to participants or communities.

Treatment and rights of research participants involving minors (Finnish National Board on Research Integrity, 2019). Although all students at both educational levels experienced the measures and interventions as a part of their regular schoolwork, their study participation was voluntary. Accordingly, study

participation or refusal did not affect students' school credits, and teachers were not aware of their students' possible refusals to participate in the study. A consent form along with an information letter was sent to students' homes few days before the first measures. The letter included information about, for example, the research aims, usage of the data, and the researchers' contact information for further questions. Students and guardians were also informed of the possibility of withdrawing from the study during any phase of the research without negative consequences. For under-aged students, their guardians signed a written consent and returned it in a closed envelope to the teacher, who delivered it to the researcher. When a signed consent was not received back, it was interpreted as negative consent.

Processing of personal data (Finnish National Board on Research Integrity, 2019). At both educational levels, students signed into the online inquiry task with a specific individual code, and students' names or Internet Protocol (IP) addresses on their computers were not stored in the systems. The data gathered in the online inquiry tasks were stored in a secured server placed in Chile (iFuCo project) and in Europe (Aroni project). After complete data collection in both research projects, the data were pseudonymized by giving students identifiers (IDs) when storing data to the secured server of the consortium universities. Physical documents, including participants' names (e.g., consents and questionnaires), were stored according to the rules of the consortium universities.

Protecting privacy in research publications and openness of the data (Finnish National Board on Research Integrity, 2019). During data analyses, to consider students' privacy and confidentiality issues, pseudonymized data were used. Further, results have been reported such that the students cannot be identified. However, the teachers received the results of their students' basic reading skills (reading fluency and comprehension tests). Sub-studies II and III have been published in open-access journals, including statements that the datasets generated and analyzed during the sub-studies are available from the corresponding author upon reasonable request.

5.3 Interventions

5.3.1 Theoretical perspectives of learning during the interventions

Both implemented interventions combined video modeling or teacher's introductions on online inquiry skills with students' independent practice and collaborative working through lessons (see Table 2). Thus, more traditional teacher-centered views of learning were accomplished with more student-centered cognitive, constructivist, and socio-constructivist approaches (cf. Greenlaw, 2015). Note that modeling videos and teachers' introductions were relatively short (5–15 minutes), and the most time was given to students' working, thinking, and discussions during the lessons. However, when the aim

is to learn new or quite challenging skills, it is useful first to model desired skills or give examples of the strategies related to the advanced skills (e.g., Coiro, 2011b), although students' responses during their practice can vary and be reflectively compared.

Phases of online inquiry, including locating, evaluating, and synthesizing information (Leu et al., 2019), formed a clear, sequenced structure for the lessons of the interventions (see Table 2). Further, the online inquiry intervention among upper secondary school students followed the principles of problem-based learning (cf. review by Yew & Goh, 2016). That is, students were given a health-related information problem to be collaboratively solved during the lessons. In the same way, the process of online inquiry (see Leu et al., 2019) shares many similarities with the broader term inquiry-based learning (IBL) (see review by Pedaste et al., 2015), such as cyclic phases of inquiry. Detailed design principles of the intervention at the upper-secondary school level are described in Kiili et al. (2022b).

5.3.2 Description of the interventions

The uniqueness of the implemented interventions (Sub-studies I and III) lies in their comprehensive processes, wherein different phases of online inquiry (Leu et al., 2019) are closely related and evaluation practices are integrated into various phases of online reading. Furthermore, in both interventions, the inquiry process was sequenced into manageable parts (cf. De Hei et al., 2016).

Both interventions were researcher-designed but implemented by class teachers (sixth grade) or language arts teachers (upper secondary school). Course-based studying limited the time available for interventions in upper secondary school more than in primary school, where the class teacher is responsible for most of the lessons during a school year. Owing to the time schedules in the research projects, the intervention group teachers at the upper secondary school level had more opportunities to influence the aims and contents of the interventions than those at the sixth-grade level. For the same reason, the intervention group teachers at the upper secondary school level received professional development for three hours. In contrast, class teachers were only introduced to each of the three modules (see Table 2) of the intervention. Table 2 summarizes the information about the intervention group teachers, their professional development, and the interventions' length, aim, content, teaching methods, and materials.

TABLE 2 Description of the Interventions in Sub-studies I and III

	Sub-study I: sixth graders	**Sub-study III: upper secondary school students**
Teachers	Primary school teachers: intervention group teachers (*N* = 8) and control group teachers (*N* = 7)	Language arts teachers: intervention group teachers (*N* = 5) and control group teachers (*N* = 6)
Teachers' professional development	A 45-min session with a researcher before each of the three modules of the intervention, including discussions of the aims, phases, and materials of the lessons	A 3-h session few months before the intervention, including theoretical insights on online inquiry strategies and an introduction to the initiative intervention plan with an opportunity to suggest modifications for it
Length of the intervention	21 x 45 min lessons	4 × 75 min lessons
Aim of the intervention	Promoting students' online inquiry skills	Promoting students' online inquiry skills
Content of the lessons	Module 1: explicit teaching and practicing of online inquiry skills—searching, evaluating, and synthesizing information (9 × 45 min) Module 2: practicing taught skills in the social science project (4 × 45 min) Module 3: practicing taught skills in the science project (8 × 45 min)	Process of online inquiry by investigating a controversial health topic: Lesson 1: searching information (75 min) Lesson 2: evaluating information (75 min) Lesson 3: synthesizing information (75 min) Lesson 4: communicating information to others (75 min)
Teaching methods	Module 1: modeling, analyzing, discussing, practicing, and reflecting online inquiry skills. Practicing taught skills in the restricted (Module 2) and open (Module 3) online environment	Teachers' short introductions on online inquiry skills, students' working in small groups, and a seminar in the last lesson. Students' working document included instructions, prompts, and guiding questions for each online inquiry skill

continues

Table 2 continues

	Sub-study I: sixth graders	Sub-study III: upper secondary school students
Materials for teachers	Package for each module, including a manual for teachers, worksheets for students, modeling videos, and Power Point slides	Microsoft OneNote digital environment, including all intervention materials (e.g., manual for teachers, task assignment, Power Point slides, and students' working documents)
Materials for students	Worksheets for each lesson	Google Docs working document, access to the task assignment, and slides in the OneNote environment.

Note. Teachers at the upper secondary school level could have taught more than one course.

Even though the main aim of the two interventions was similar—promoting students' abilities to search, evaluate, and synthesize online information—the contents of the lessons were adjusted to the students' age. For upper secondary school students (Sub-study III), the topics of the sub-tasks during lessons were more complicated, and the sub-tasks required higher-level cognitive abilities than those of sixth graders (Sub-study I). However, in both interventions, contrasting cases and/or controversial online texts were applied (cf. Braasch et al., 2013). Further, upper secondary school students freely selected authentic online texts from the Internet during their online inquiry. In contrast, sixth graders processed also pre-designed and pre-selected online texts.

Teaching methods were also selected according to students' age and cognitive skills. Thus, older students (Sub-study III) were responsible for their learning process, whereas younger students' lessons (Sub-study I) were more teacher-centered by nature. Accordingly, upper secondary school students collaboratively practiced skills through online inquiry (4 × 75 min). During the learning process, the teacher briefly introduced each online inquiry skill, and the prompts in the working document scaffolded students' work (searching for, critically evaluating, and synthesizing information). In the seminar during the last lesson, students presented and shared their learning in small groups. Therefore, the online inquiry process required students' self-monitoring (see Stadtler & Bromme, 2007) and abilities to share ideas and work collaboratively (e.g., Kiili et al., 2019).

For sixth graders (Sub-study I), each online inquiry skill was first briefly modeled with a video. While watching the video, students individually analyzed the modeled skill by responding to the questions in the worksheet. Next, they discussed the responses with their peers and the teacher. Video modeling was used to motivate students' learning (cf. Choi & Johnson, 2005). During the following lessons, students practiced each taught skill with different sub-tasks. For example, in the evaluation lessons, they evaluated the credibility of two online texts with worksheets. Paper versions of the online texts were used to

maintain younger students' attention in learning a specific online inquiry skill, as computers could hamper their concentration. In the following projects, sixth graders practiced taught skills first in a restricted environment and then in an open online environment. Thus, teaching and learning online inquiry skills through modules progressively moved from more manageable sub-tasks to more difficult ones (cf. Sparks et al., 2021).

As the interventions were part of regular schoolwork, teachers could decide how they personally evaluated students' work. Some teachers at the upper secondary school level gave numeric credit for students' collaborative work (working documents), but not all of them. Therefore, these credits were not applied as data in this dissertation.

5.3.3 Fidelity of the interventions

The fidelity (McKenna et al., 2014) of both interventions was ensured in several ways. Before each of the three intervention modules among sixth graders (Sub-study I), the researcher introduced the aims, materials, and assignments to the intervention group teachers. As the language arts teachers at the upper secondary school level (Sub-study III) were more involved in planning the intervention than the primary school teachers, the final materials were shared with them via OneNote digital environment. The intervention group teachers also received a manual and time schedules for each lesson at both educational levels.

During the interventions, the intervention group teachers were tasked with marking any deviations from the intervention plan in their diaries. They were also able to contact the researcher to ask questions. Further, the researchers followed part of the lessons during both interventions. In Sub-study III, the control group teachers also reported how often they taught online inquiry skills during their regular Texts and Influence course. After the interventions, the researchers collected the worksheets (Sub-study I) and working documents (Sub-study III) filled by students. At the upper secondary school level, all the intervention group teachers were interviewed.

5.4 Measures and data collection

5.4.1 Online inquiry tasks measuring critical online reading skills

In all sub-studies, students' critical online reading skills were measured with online inquiry tasks. In Sub-studies I and III, online inquiry tasks were used as pre- and post-tests, whereas in Sub-study II, only data from the pre-test were applied. The performance-based tasks at both the primary and the upper secondary school levels followed the online reading and comprehension model (Leu et al., 2019), including the phases of searching, selecting, analyzing, evaluating, and synthesizing online information. In addition, upper secondary school students' task included a first phase where students were asked to specify their information need (see Table 3). Furthermore, tasks were adjusted to students' age level, for example, using different topics, task assignments, prompts, and timing at the sixth grade and upper secondary school levels (see Table 3). In the following sections, these features of the online inquiry tasks are described in more detail, and the tasks for different age levels are compared.

TABLE 3 Online Inquiry Tasks in Sub-studies I–III

	Sub-study I: sixth graders	**Sub-studies II and III: upper secondary school students**
Topics	Computer gaming and reading on screen	Vaccination and fats
Task scenarios	Students were asked to explore the topic on the Internet to write an article for a school magazine (computer gaming topic) OR an email message for a student council (reading on screen topic) on the advantages and disadvantages of the topic.	Students were asked to explore the topic on the Internet to help a fictitious expectant mother to decide whether to vaccinate her child (vaccination topic) OR to help a fictitious student to decide whether to avoid saturated fats in their diet (fats topic).

continues

Table 3 continues

	Sub-study I: sixth graders	Sub-studies II and III: upper secondary school students
Task prompts	1. Use the search engine to search for the most credible online texts that you can later apply in your article/email message. Select the three most useful texts and save them by pressing the "Select the page" button. (max. 8 min) 2. Use the snippet tool to mark the two most important sentences in each text. The selected snippets will be saved for your later use. Snippets cannot be longer than 20 words. (max. 12 min) 3. How credible is this online text? How many stars do you give it? (stars 1–5) Why do you think so? (open question) (max. 7 min) 4. Write your article/email message here. It should be at least 50 words. Remember to use your own words. Do not copy from the snippets. (max. 15 min)	1. What kind of information do you need to advice the expectant mother OR the student? (open question) 2. Search for and select three online texts that help you to provide credible information (copy of URL addresses). 3. Which are the three main ideas in the text that you can utilize in your response? What aspects make the online text credible? What aspects may weaken the credibility of the online text? (open questions) 4. What is your position on whether the expectant mother should vaccinate her child OR whether the student should avoid saturated fats in their diet? (multiple-choice question) Write below the justifications that support your position. Indicate the sources you rely on.
Task environment	Web-based environment Neurone, including a closed search engine, task prompts, and a tutorial for using the features of the environment	Web-based environment, including Google custom search engine, task prompts, and instructions
Texts in search engines	17 online texts per topic On both topics, there were researcher-designed texts (n = 3) that varied by - author - type - perspective - position on the topic Further, the links of the authentic texts (n = 14) included keywords that appeared in the task scenarios, but the texts concerned issues that were not relevant to the task at hand.	35 authentic online texts per topic Texts varied according to their usefulness (relevance and credibility) for the task (see McCrudden, 2018). On both topics, there were - most useful texts (n = 3) - useful texts (n = 5) - less useful texts (n = 5) - not useful texts (n = 22)
Timing	Max. 42 min for the entire task	Max. 60 min for the entire task

Web-based task environments. Both Web-based task environments were designed for research purposes and comprised task prompts, instructions, and a customized search engine with preselected online texts. Sixth graders' Web-based task environment, titled Neurone (González-Ibañez et al., 2017; Sormunen et al., 2018), included a digital tutorial to help students navigate and proceed through the task phases (Sub-study I). To motivate sixth graders, two virtual students guided them in proceeding in the system and gave them all task and sub-task assignments (see also Kullberg et al., 2023). For upper secondary school students (Sub-studies II and III), written instructions were displayed next to the task prompts.

Students' responses and the time used for completing the tasks were recorded in both task environments. The task timing was also shown to students. For sixth graders, time was restricted in each task phase (see Table 3), and they received a reminder three minutes before the time ended. The time limit aimed to help students reasonably share their working time between different task phases during a 45-min lesson. In upper secondary school students' tasks (Sub-studies II and III), reasonable time periods for each task phase were suggested by color in the time frame. However, students could use as much time as they wanted for each task phase, provided that the total time of 60 min was not exceeded. In addition, upper secondary school students could move between the different task phases using forward and backward buttons when working on their task. However, they could not change their responses after leaving a task phase (see also Kiili et al., 2021).

Task scenarios. At the beginning of the online inquiry task, students were presented with a real-life problem as a task scenario (cf. Kammerer et al., 2015; Scharrer et al., 2019). In all sub-studies, students investigated controversial topics, including their benefits and harmful aspects (see Table 3), because controversies in documents have been shown to promote students' evaluations and sourcing (e.g., Kammerer et al., 2016b; Stadtler & Bromme, 2014). Therefore, in all task scenarios, controversy was also highlighted. Accordingly, half of the sixth graders were tasked with writing a newspaper article regarding the advantages and disadvantages of computer gaming, and the remaining half were tasked with writing an email regarding the advantages and disadvantages of reading on screen. These topics were chosen owing to their relevance to students' lives, and all students were assumed to be somewhat familiar with both. Similarly, two task scenarios were provided for upper secondary school students (Sub-studies II and III). Half of the students were tasked with helping a fictitious expectant mother to decide whether she should vaccinate her child (vaccination topic). The remaining half of the upper secondary school students were assigned to help a fictional student determine whether they should avoid saturated fats in their diet (fats topic).

Searching and selecting phase. Customized search engines were used in the searching phases, as they not only enabled the use of authentic online texts but also allowed some control over students' text selections. Table 3 briefly describes the online texts included in the search engines of the sub-studies. For

sixth graders (Sub-study I), 17 online texts were available, including 3 researcher-designed texts and 14 authentic online texts per topic. The authentic irrelevant texts included keywords that appeared in the task assignment (e.g., computer gaming) but addressed irrelevant issues. Instead, the three researcher-designed texts were supposed to be the most useful for completing the task. Respectively, 35 authentic online texts were available for upper secondary school students (Sub-studies II and III), which varied in relevance and credibility (see McCrudden, 2018). The most useful texts had the highest source credibility and text relevance for the task at hand. The smaller number of texts made the searching and selection phases easier for younger students. In addition, the system gave all sixth graders the same three researcher-designed texts after 8 min of search and selection attempts. Thus, sixth graders proceeded with the same online texts through the subsequent task phases (evaluating, analyzing, and synthesizing).

Analyzing and evaluating phase. In the next phase, all students were tasked with identifying the main ideas in each online text (see Table 3). Sixth graders were given a digital snippet tool, which they used to select three relevant ideas from each online text (see Kullberg et al., 2023), whereas upper secondary school students wrote down the three main ideas in each text. Next, students were asked to evaluate the credibility of the online texts. Sixth graders (Sub-study I) first rated the credibility of each text with stars ranging from 1 to 5. Stars given made writing justifications for credibility more effortless ("Why do you think so?"). Upper secondary school students (Sub-studies II and III) were tasked with considering the strengthening and weakening aspects of each text's credibility. They were asked using separate open-ended questions, as previous research has shown that confirming the credibility of online texts requires different abilities than questioning the credibility (e.g., Kiili et al., 2018a). The latter also seems to be a more demanding skill for students (e.g., Kiili et al., 2023).

Synthesizing phase. In the final phase of both online inquiry tasks (see Table 3), students composed a written product (synthesis). Instructions emphasized using three given (sixth graders) or self-selected (upper secondary school students) online texts in the written product but not copying from them. To help younger students (Sub-study I) write, they were given a title for their essay: "Advantages and disadvantages of computer gaming/reading on screen." Older students (Sub-studies II and III) selected their position on the topic and justified it by writing an essay. While writing, sixth graders could see their selected snippets, and by clicking on them, they could read each online text. For upper secondary school students, the online texts and their responses in earlier task phases were available.

5.4.2 Internet-specific epistemic justifications

In Sub-study II, when examining upper secondary school students' credibility evaluation skills before the intervention, the ISEJ inventory developed by Bråten et al. (2019a) was applied to measure students' beliefs in their justifications for knowing in the Internet context. Previous research has shown that older students'

Internet-specific justifications for knowing can contribute to their online reading skills (e.g., Kammerer et al., 2013, 2021; see also, intervention study by Bråten et al., 2022). The original measure (Bråten et al., 2019a) was translated into Finnish and adapted for upper secondary school students.

The inventory included three knowing dimensions, which were all measured with four items: 1) personal justification (e.g., "When I find information on the Internet, I evaluate whether this information is consistent with my own understanding of the topic."), 2) justification by authority (e.g., "To determine whether the information I find on the Internet is trustworthy, I evaluate whether the author has sufficient knowledge of the topic."), and 3) justification by multiple sources (e.g., "I evaluate the claims I find on the Internet by checking several information sources on the same topic."). A five-point Likert scale was applied (from 1 = *strongly disagree* to 5 = *strongly agree*) with all options expressed in words.

In this dissertation, students' epistemic beliefs were also regarded as individual difference factors, the role of which in students' critical online reading skills was investigated as part of RQ3.

5.4.3 Other measures

As students' basic reading skills create a foundation for their critical online reading skills (cf. Coiro, 2011a), students' reading fluency was measured in all sub-studies. A word chain test by Holopainen et al. (2004) was used at the primary and upper secondary school levels, including 25 chains with 4 words written with no spaces in between. Within 90 s, students were asked to separate as many chains into primary words as possible; thus, their scores varied between 0 and 100.

Critical reading also has roots in reading comprehension skills (e.g., Christodoulou & Diakidoy, 2020), as without being able to understand the content of the text, critical reading cannot be employed. Therefore, sixth graders (Sub-study I) completed a reading comprehension test (Kajamies, 2017) in which they read a text concerning the diversity of nature and answered, for example, three open-ended questions on its main ideas. As the maximum score for each question was 6 points, students' total scores varied between 0 and 18. The following Kappa values (Cohen, 1960) were obtained for inter-rater reliability between two researchers (20% of the responses scored): 0.90 (Question 1), 0.68 (Question 2), and 0.95 (Question 3).

Because topic knowledge affects reading comprehension (e.g., Cervetti & Wright, 2020; Kintsch, 1988) and may play a role in critical online reading skills (Anmarkrud et al., 2021), upper secondary school students' prior knowledge about the topic was measured in Sub-studies II and III. The measure comprised ten statements on vaccination or fats: three correct and seven incorrect. The students were asked to select three statements they considered to be the correct ones. They earned one point for each accurate or non-selected incorrect statement (0 or 1 per statement). Six items in both topics were included in students' final

scores (0–6). Reliability for vaccination was 0.82 with a 95% *CI* [0.68–0.96] and for fats was 0.94 with a 95% *CI* [0.91–0.96] (Raykov et al., 2010).

The variables mentioned above served as control variables in the sub-studies. Furthermore, in this dissertation, RQ3 investigated the role of these individual differences in students' critical online reading skills.

5.4.4 Data collection procedures

The data of the sub-studies were collected during regular schoolwork. Students completed the online inquiry tasks during a 45-min lesson (sixth graders) or 75-min lesson (upper secondary school students) in classrooms. Upper secondary school students filled in the ISEJ inventory (Sub-study II) before the task and returned it to the teacher. In addition, they responded to a reading fluency test just before the first online inquiry task (Sub-studies II and III). In contrast, sixth graders (Sub-study I) completed both reading tests in a lesson a week before the first online inquiry task.

Among sixth graders, the entire class had the same task topic (computer gaming or reading on screen). In contrast, a researcher randomly allocated the vaccination topic to half of the upper secondary school students and the fats topic to the remaining half in each course. Each student's topic in the post-test (Sub-studies I and III) differed from their pre-test topic. Upper secondary school students responded to the prior topic knowledge test just before the online inquiry task on the corresponding topic (pre- and post-test). Further, sixth graders completed post-tests a week or two after the last intervention lesson. In contrast, upper secondary school students completed the post-test in the lesson after the last intervention lesson. The researcher gave instructions to the students and helped them if they encountered technical problems during the online inquiry task.

5.5 Data analysis

5.5.1 Analyses of students' responses in an online inquiry task

As Table 4 shows, sixth graders' qualitative data included written justifications for the credibility of online texts and their composed written products (Sub-study I). Respectively, qualitative data for Sub-study II comprised upper secondary school students' written justifications for the credibility of online texts. As they selected different online texts, their text selections were categorized, scored, and applied as a control variable in Sub-study II. In contrast, qualitative data for Sub-study III covered several phases of the online inquiry task, including specified information need, formulated search queries, written credibility judgments, and composed written products (see Table 4).

TABLE 4 Data, Scoring, and Variables Based on Students' Responses in Different Phases of the Online Inquiry Task

		Phase of online inquiry task			
		Specifying information need	Searching for and selecting information	Evaluating the credibility of information	Composing a written product
Sub-study I	Data			1026 written justifications in pre- and post-tests	342 written products in pre- and post-tests
	Scoring			Number of relevant justifications for credibility across three online texts	Number of justifications for credibility used in the written product
	Variables			Expertise of the source, other source features, argumentation in the text, and other aspects of the content	Source features and quality of the content
Sub-study II	Data		1031* text selections	1035 justifications for credibility regarding the strengthening AND weakening aspects	
	Scoring		Usefulness (relevance and credibility) of the three self-selected online texts	Five credibility aspects acknowledged in justifications: author, venue, intentions, evidence, and corroboration AND justifications at the highest level of reasoning across three self-selected online texts	
	Variables		Selection score (0–9)	Evaluation performance (0–5)	

continues

Table 4 continues

		Phase of online inquiry task			
		Specifying information need	Searching for and selecting information	Evaluating the credibility of information	Composing a written product
Sub-study III	Data	365 written responses in pre- and post-tests	365 search queries in pre-and post-tests	1095 justifications for credibility regarding strengthening AND weakening aspects in pre- and post-tests	365 written products in pre- and post-tests
	Scoring	Use of source features or evaluative comments in response	Number of source features applied in search queries	Evaluation of source features (author, venue, or intentions) in responses across the three self-selected online texts	Sources mentioned, source–source links, source–content links, and evaluative comments used in the written product
	Variables	Sourcing in specifying information need (0–3)	Sourcing in search queries	Sourcing in credibility judgments (0–7)	Sourcing in written product (0–5)

Data analysis in Sub-study I. In the pre- and post-tests, sixth graders justified the credibility of each of the three researcher-designed online texts by answering the open question: "Why do you think so? (stars given for the credibility of the text). As writing justifications is a challenging task for sixth graders, all relevant justifications related to source information (e.g., Bråten et al., 2018b; Britt & Aglinskas, 2002) or the quality of the content (e.g., Braasch et al., 2013; Britt et al., 2014) were identified through qualitative content analysis (e.g., White & Marsh, 2006). Credibility aspects based on the above-mentioned previous studies were deductively searched from students' responses, but the content analysis process was also inductive and data-driven. One written sentence could include one or more justifications for credibility.

Four main categories were formed based on the identification of justifications in students' responses. Two were related to source information and titled: expertise of the source (e.g., author, publisher) and other source features (e.g., date and type of online text, availability of contact information). Respectively, two of them were related to the quality of the content and titled: argumentation in the text (e.g., research basis, quality of evidence, consideration of both sides of an issue) and other aspects of the content (e.g., correspondence with one's own prior knowledge and experiences, the writing style of the online text). After categorization, four count variables were established based on the total number of justifications across the three evaluated

online texts. The reliability of the categorization was calculated for 15% of the students' justifications. The inter-rater agreement for the categorization was 0.90 (Cohen's kappa; Cohen, 1960).

To explore how sixth graders' justifications for the credibility of online texts were reflected in their written products, whether their justification or part of it (e.g., author, publisher, or quality of evidence) appeared in their writings was examined. The found overlap was categorized as representing one of the main categories of the justifications for credibility: source features or quality of content. Based on categorization, two count variables were formed.

Data analysis in Sub-study II. Upper secondary school students justified the credibility of each of the three self-selected online texts by answering two open questions: What aspects make the text credible? What aspects may weaken the credibility of the text? Data for Sub-study II were derived only from the pre-test. As students mentioned only a few relevant weakening aspects per text, these two responses were considered one unit of analysis for each self-selected online text. The qualitative content analysis (e.g., White & Marsh, 2006) of students' responses deductively proceeded based on earlier research that highlights the central aspects of credibility. Accordingly, students' written justifications that were related to five credibility aspects—author, venue, intentions (e.g., Bråten et al., 2018b), evidence (e.g., Forzani, 2020; Sinatra & Lombardi, 2020), and corroboration (e.g., Kohnen & Mertens, 2019; Wineburg, 1991)—were identified. Three of them (author, venue, and intentions) can be regarded as source information, whereas evidence and corroboration refer to the content of the text. Similar to sixth graders' responses (Sub-study I), one written sentence could include one or more justifications for credibility.

In Sub-study II, the aim was to not only count the number of students' relevant justifications but also analyze the depth of their reasoning (see also Coiro et al., 2015; Kiili et al., 2019), as older students should reach a more sophisticated level in their evaluations than sixth graders. Thus, students' justifications for each of the five aspects were categorized at four levels, from 0 (student does not refer to the evaluation criterion) to 3 (student engages in deep reasoning when evaluating the credibility aspect). The final evaluation score (0–5) was based on the different credibility aspects acknowledged and the depth in students' reasoning across the three self-selected online texts. For inter-rater reliability, the Kappa value (Cohen, 1960) was calculated for each aspect scored by two researchers (10% of responses) and varied from 0.78 to 0.90.

Further, upper secondary school students' text selections were analyzed in Sub-study II. During the online inquiry task, students could select three authentic online texts regarding either vaccination or fats. The selected texts were scored from 0 to 3 according to their usefulness (text relevance and source credibility; see McCrudden, 2018). Students' score for text selections (0–9) were received by counting together the scores of the three online texts.

Data analysis in Sub-study III. The qualitative data for Sub-study III comprised upper secondary school students' written responses to the open questions regarding specifying their information need and evaluating the credibility of online

texts (see Sub-study II) in pre- and post-tests. Furthermore, students' search queries and written products were analyzed from the pre- and post-tests.

Scoring rubrics for sourcing variables in Sub-study III were developed in a study by Kiili et al. (2021). However, in their study, sourcing in search queries was scored 0–3, but in Sub-study III, it was a continuous count variable. Further, students' evaluations of credibility aspects in Sub-study II formed the basis for variable "sourcing in credibility judgments" in Sub-study III. Accordingly, only students' justifications for credibility aspects reflecting source information (author, venue, and intentions) were applied. Three of the four sourcing variables (sourcing in specifying information need, search queries, and credibility judgments) were scored according to the number of source information and evaluative comments in students' responses. In addition, in variable sourcing in written product, the quality of students' use of source information (source–content links, source–source links, and/or evaluative statements applied) was taken into account. Kappa values (Cohen, 1960) for inter-rater reliability between two researchers (10% of the responses scored) varied from 0.75 to 0.92 in the four sourcing variables.

5.5.2 Statistical analyses

In Sub-studies I and III, linear and negative binomial (see Coxe et al., 2009) regression analyses were used to measure the efficacy of the interventions. These analyses enabled controlling important variables (e.g., students' pre-test scores, basic reading skills, prior topic knowledge, and topic order in pre- and post-tests), which may affect students' learning of critical online reading skills during the intervention. In addition, when parameters were estimated with Mplus software (Muthén & Muthén, 1998–2017), for example, the non-normality of some variables, missing data, and intra-class correlations in the data could be considered. Further, Wilcoxon's test was applied in investigating whether intervention enhanced sixth graders' use of justifications for credibility in their written products (Sub-study I).

By counting the reliable change index (RCI; Jacobson & Truax, 1991), the number of students whose sourcing performance changed, either negatively or positively, and those whose performance did not change during the intervention were determined (Sub-study III). In addition, the roles of control variables in students' RCIs were investigated by bootstrap analysis with 95% *CI*s for mean differences (Efron, 1987) and crosstabulation with the χ^2 test.

In Sub-study II, confirmatory factor analysis was used to investigate whether data confirmed the original three-dimensional structure of the ISEJ inventory. Next, a hierarchical regression analysis within the structural equation modeling (SEM) framework (de Jong & van der Leij, 1999) was used to examine the unique effects of ISEJ dimensions on students' evaluation performance when students' text selections, reading fluency, and prior topic knowledge were controlled for. Finally, topic differences in the linkages between ISEJ dimensions and students' evaluation performance were examined using the multi-group procedure.

6 MAIN RESULTS OF THE SUB-STUDIES

This dissertation explored students' critical online reading skills and the factors associated with their skills through Sub-studies I–III. In addition, Sub-studies I and III investigated the effects of an intervention. This chapter summarizes the results of the sub-studies. Terminology is consistently used across the studies, differing to some extent from the terms used in the original articles to make comparing the studies easier.

6.1 Sub-study I: Promoting sixth graders' credibility evaluation of Web pages: An intervention study

Sub-study I investigated how sixth graders justified the credibility of online texts and whether an intervention increased their relevant justifications. Further, it examined how students' justifications for the credibility of online texts were reflected in their written product and whether these reflections increased during the intervention.

Sixth graders from ten primary schools in Finland composed an intervention group (190 students) and a control group (152 students). Pre- and post-tests comprised online inquiry tasks in which students searched, selected, evaluated, and synthesized information. Sixth graders evaluated three online texts on computer gaming or reading on screen. They rated the credibility of each text with stars (1–5) and justified their ratings by answering the question, "Why do you think so?" At the end of the online inquiry task, students composed a short, written product, which was supposed to synthesize the benefits and harmful effects of the topic.

Between the pre- and post-tests, the class teachers implemented the intervention program (21 × 45 min lessons) during a six-week course as part of regular schoolwork. First, the intervention group students received explicit instruction on critical online reading skills: searching for information, evaluating the credibility of information, and synthesizing information. Further, they

practiced the explicitly taught skills in two online reading projects. In the first credibility evaluation lesson, the intervention group students watched a short video in which a more advanced and a less advanced virtual student modeled the evaluation of a newspaper article. Then, by responding to questions in the worksheet, they analyzed the virtual students' evaluation strategies. Next, they shared responses with a peer and during a discussion led by teacher. In the following lessons, students practiced evaluating two controversial online texts with a worksheet including questions about, for example, the authors and their expertise, intentions, main ideas, and overall credibility of each text. Finally, students' responses were shared and reflected with the teacher.

Results showed that in the pre-tests, students most often justified the credibility of texts by referencing the expertise of the source (M = 1.56; SD = 1.95). However, variation was considerable, indicating that some students presented many justifications related to this aspect, whereas some students did not present any. In contrast, students rarely referred to other credibility aspects in their justifications: other source features (M = 0.52; SD = 1.05), argumentation in the text (M = 0.33; SD = 0.71), and other aspects of the content (M = 0.39; SD = 0.79). Further, only 8% of the students used their justifications for credibility in their written product.

Background variables (e.g., pre-test scores, reading comprehension, reading fluency, gender, and topic order) were controlled for in multilevel negative binomial regression analysis. Based on the results, the explicit teaching of online inquiry skills and practicing of those skills in two projects enhanced sixth graders' justifications for the credibility of online texts by referencing source information. After the intervention, students in the intervention group justified their credibility ratings 1.52 times more often with the expertise of the source and 1.83 times more often with other source features than students in the control group. For this dissertation, effect sizes were calculated (Coxe, 2018): standardized mean difference (SMD) = 0.32 [95% CI 0.09–0.66] for the expertise of the source and SMD = 0.39 [95% CI 0.13–0.82] for other source features, indicating small effects (Cohen, 1988).

In contrast, sixth graders' justifications for credibility related to argumentation in the text or other aspects of the content did not significantly increase. In addition, about 11% of the sixth graders used justifications for credibility in their post-test essays, suggesting that the intervention did not enhance students' use of justifications in their written products.

Further, all students' pre-test scores were associated with their post-test scores in the corresponding category. In the post-tests, the better reading comprehension skills sixth graders had, the more often they paid attention to the expertise of the source and argumentation in the text, and vice versa. In addition, students who completed the computer gaming task in the post-test scored better in justifying the credibility by referencing argumentation in the text than those who completed the reading on screen task in the post-test. Students' reading fluency and gender were not associated with the number of relevant justifications in the post-tests.

To conclude, the intervention (21 × 45 min lessons), including the explicit teaching of online inquiry skills and practicing of these skills in two projects, succeeded in promoting sixth graders' attention to and evaluation of the source information but not the content of online texts. In all, students' performance level remained low during the intervention, suggesting that most sixth graders had difficulties attending to the credibility aspects of online texts, writing justifications for credibility, and utilizing justifications for credibility in their written product. In addition, students' pre-intervention evaluation skills, reading comprehension, and topic order predicted their evaluation performance after the intervention.

6.2 Sub-study II: Students' abilities to evaluate the credibility of online texts: Role of Internet-specific epistemic justifications

Sub-study II aimed to investigate upper secondary school students' abilities to justify the credibility of self-selected online texts during online inquiry. Further, associations between students' ISEJ and their justifications for the credibility of online texts were examined.

Before an online inquiry task, students completed the ISEJ inventory. The inventory included three dimensions reflecting how students believed they evaluate online texts during reading: personal justification, justification by authority, and justification by multiple sources. In the online inquiry task, students were asked to explore either of two health topics: vaccination or fats. With the Google custom search engine, students selected three authentic online texts. Further, they answered the questions: "What aspects make the text credible?" "What aspects may weaken the credibility of the text?"

Results of Sub-study II suggested considerable differences in upper secondary school students' abilities to justify the credibility of online texts. Students most often paid attention to the venue and evidence presented in the online texts. In particular, almost 90% of the students evaluated the venue, and over 75% evaluated the evidence at least once across three online texts. In contrast, less than 29% of the students justified the credibility of online texts by considering intentions at least once, and less than 14% referred to corroboration at least once across the three online texts.

Students most often reached the highest level in their reasoning when they evaluated evidence or venue. More than 26% and 20% of the students evaluated evidence and venue, respectively, at least once at the highest level across the three online texts. However, students very rarely reached a deep level in their reasoning when evaluating the intentions (less than 7% of the students) and referring to corroboration (less than 3%).

On average, students scored 3.07 (0–5) for their evaluation performance. Over 37% of the students demonstrated highly versatile ability to justify the credibility of online texts. However, almost 10% of the students performed very

poorly. An additional 20% of the students demonstrated having limited evaluation skills. Students who explored fats scored statistically significantly higher (3.22, *SD* = 1.02) than students who examined vaccination (2.90, *SD* = 1.16).

When the associations between students' epistemic beliefs about how they evaluate online texts during reading and their evaluation performance were examined, students' text selections, reading fluency, and prior topic knowledge were controlled for. The more students believed that they evaluated authority or compared multiple sources when reading online texts, the better their evaluation performance was. Students' beliefs about personal justification were not associated with their evaluation performance.

Further, ISEJ similarly explained students' evaluation performance in both topics, although students exploring the fats topic performed better in prior topic knowledge tests and in selecting and evaluating online texts than students investigating the vaccination topic. In addition, the better students were at selecting useful online texts and the better reading fluency they had, the better they performed in justifying the credibility of online texts.

In sum, the study revealed remarkable differences in upper secondary school students' credibility evaluation skills. The differences were related to the credibility aspects of the online texts students attended to and the depth of their reasoning. Note that one-tenth of the students performed exceptionally poorly in justifying the credibility of online texts. Students' epistemic beliefs about evaluating the authority and corroborating information during reading online explained their evaluation performance. Further, students' text selections and reading fluency were associated with their abilities to justify the credibility of online texts.

6.3 Sub-study III: Teaching sourcing during online inquiry—adolescents with the weakest skills benefited the most

In Sub-study III, the efficacy of an intervention aimed to foster upper secondary school students' sourcing during online inquiry was investigated. In addition, the study examined how students' sourcing skills changed during the intervention and how students' pre-intervention sourcing skills, reading fluency, prior topic knowledge, and topic order in tasks were associated with the changes.

The same students from eight Finnish upper secondary schools as in Sub-study II were divided into two conditions: an intervention group (196 students) and a control group (169 students). Between the pre- and post-tests, the intervention group participated in an intervention (4 × 75 min lessons) on online inquiry as part of their Texts and Influence course. During the intervention, students investigated one of four controversial health topics in small groups of 2–4 students. The lessons were based on phases of online inquiry: searching for information, evaluating information, synthesizing information, and communicating information to others. In the first lesson, students received a task

assignment, selected their topic, and formed small groups. During the first three lessons, the teacher briefly introduced the students to a specific online inquiry skill (searching, evaluating, or synthesizing online information). After each introduction, students worked in small groups for the rest of the lesson. Students' work was supported by Google Docs document, which included guiding and reflection prompts for each online inquiry skill. Finally, in the seminar during the fourth lesson, students shared their responses and learning.

In the pre- and post-tests, students completed the same online inquiry task as in Sub-study II. In the first phase of the task, students were asked to specify their information need by responding to an open-ended question. Next, they searched for information, and their search queries were recorded. After the selection and evaluation phases (see Sub-study II), students wrote justifications for their position on the topic based on the online texts they had previously selected and evaluated.

In the pre-test, upper secondary school students did not perform extremely well in sourcing during different phases of online inquiry. Only few students included multiple source features or evaluative comments in their responses when they specified their information need and applied several source features (e.g., organizations, credentials, names of persons relevant to the topic, and type of the document) across their search queries. Moreover, when students justified the credibility of online texts or composed a written product, more variation in sourcing was found among students. Accordingly, some students justified the credibility of online texts by referencing different source features (author, venue, intentions) and applied several source–content links, source–source links, and evaluative comments in their written product, whereas for some students, engaging in sourcing was not a noteworthy practice during these phases of online inquiry.

After the background variables (pre-test scores, topic order, prior topic knowledge, and reading fluency) were controlled for, the results showed that the explicit teaching of online inquiry skills and students' collaborative work in small groups fostered students' sourcing in search queries, credibility judgments, and written product compared with the controls. Effect sizes (Cohen's d) were 0.39 [95% *CI* 0.17–0.61] for sourcing in credibility judgments and 0.37 [95% *CI* 0.15–0.58] for sourcing in the written product, and *SMD* = 0.14 [95% *CI* 0.03–0.41] for sourcing in search queries, indicating small effects (Cohen, 1988). Further, students' sourcing in specifying information need was not improved.

Although the skills of only a limited number of students (4%–25%) were improved, RCI analysis revealed that the intervention significantly enhanced the sourcing skills of the worst-performing students in the pre-test. Further, topic order in pre- and post-tests predicted some changes in students' sourcing skills, indicating that students used sourcing more often when they investigated the vaccination topic than when they explored the fats topic. Students' reading fluency and prior topic knowledge were not associated with any changes in students' sourcing skills during the intervention.

To conclude, the intervention (4 × 75 min lessons), including the explicit teaching of online inquiry skills and collaborative working in small groups, promoted upper secondary school students' sourcing skills. However, students did not increase their sourcing in specifying information need. Most importantly, students with the weakest skills benefited the most from the intervention, including teachers' short introductions and collaborative working through online inquiry. Further, topic order explained some changes in students' sourcing skills. Overall, upper secondary school students could have achieved better sourcing skills during the intervention.

7 DISCUSSION OF MAIN FINDINGS

7.1 Students' critical online reading skills considerably varied

The first research question of this dissertation aimed to examine students' abilities to justify the credibility of online texts and engage in sourcing. First, the results confirmed findings from previous research (e.g., Forzani, 2022; Kiili et al., 2008, 2018b), indicating a considerable variation in students' skills in both age groups. Among sixth graders and upper secondary school students, some students possessed high-level critical online reading skills, but others did not. It has been suggested that critical evaluation skills develop stepwise along with adolescents' maturation (cf. Potocki et al., 2020). Nevertheless, evidence in this dissertation suggests that this process can significantly vary among students (see also Sparks et al., 2021). Some students at primary school can learn and apply quite sophisticated evaluation strategies, whereas some students at the upper secondary school may need to learn basic abilities.

Table 5 presents examples of students' justifications for the credibility of online texts at the primary (Sub-study I) and upper secondary school levels (Sub-study II). As seen in the examples, sixth graders' responses mainly differed in how many credibility aspects (e.g., author, venue, evidence) they paid attention to in the online texts. In contrast, upper secondary school students' responses also differed in the level of their reasoning. Accordingly, some students' justifications were thorough, whereas others' reasoning was relatively superficial.

TABLE 5 Examples of Students' Justifications for the Credibility of Online Texts in Sub-studies I and II

	Examples of students' justifications		
Sub-study I: sixth graders	The online text seems to be credible. (ID 1216)	I think this online text is credible because it includes the ideas of a pediatrician. (ID 3341)	The text was written by a pediatrician, who had noticed how computer gaming is harmful for young children but provides the benefits of gaming too. The text was also quite new. (ID 3116)
	It sounds true. (ID 2042)	I assume that this online text is quite credible because its style of writing is correct. (ID 3308)	I thought so because the name of the author was mentioned, and it was a research-based article, which was written only a year ago. It also included names of the universities and which things have been studied there. (ID 1115)
Sub-study II: upper secondary school students	The online text was the first in search page results and the text makes a credible expression. (ID 2136)	Websites of Duodecim are used by doctors; therefore, there is only information that is based on truth. (ID 1150)	The publisher of the online text is THL, Finnish Institute for Health and Welfare, which investigates issues related to public health. Employers in THL are experts who operate in the administrative branch of the Ministry of Social Affairs. (ID 1149)
	There is lot of information, no commercials and websites are official. (ID 1287)	The author is the doctor in medicine who has used and analyzed study results in his text. There is also a list of references. (ID 1226)	The online text is written by Antti Aro, who is a professor and specialized in internal medicine. Thus, he is highly educated and an expert in the field in question. (ID 2108)

Furthermore, many sixth graders did not give any relevant justifications for the credibility of the three online texts (Sub-study I). Similarly, only a few sixth graders utilized justifications for credibility in their written products. Previous studies (e.g., Coiro et al., 2015; Forzani, 2018) have also shown that critical online reading skills are complex for young adolescents. The challenges are most apparent when an evaluation task requires students to write their responses (cf. Coiro et al., 2015; also Anmarkrud et al., 2021). Accordingly, writing justifications

for the credibility of online texts is more demanding than, for example, displaying credibility rankings for texts (cf. Zhang & Duke, 2011).

Thus, many sixth graders probably could not apply their shallow skills by writing their responses (cf. Paul et al., 2017). Also, Sparks et al. (2021) found that the evaluation task requiring written justifications for information quantity and accuracy by comparing one text to another credible text, which was supposed to be the easiest task in their study, was surprisingly difficult for some adolescents. However, separate prompts to justify each credibility aspect at the time would probably make the justification task easier for sixth graders than justifying the credibility of the online text in one question (cf. Kiili et al., 2023).

The justifications for the credibility of online texts by some sixth graders were as well written as those by upper secondary school students on average (see Table 5). A few sixth graders also took advantage of their evaluations in their written product. Thus, high variation between students' skills already exists at the lower school levels. In Sub-study I, remarkable differences between sixth graders' critical online reading skills were found almost in each school class, and students' basic reading skills, for example, did not explain all of these differences. Interesting and rarely studied questions for future are how and when these differences develop (cf. Potocki et al., 2020). Factors outside the school, such as parents' own critical online reading skills, discussions within the family and with peers, or the literacy environment at home, may affect young students' abilities to read critically. Thus, to diminish developing differences in young students' critical online reading skills, teaching interventions should be implemented sufficiently early.

Second, students differently paid attention to the aspects of credibility in their justifications (see Table 5). Sixth graders considerably more often referenced source information than content (Sub-study I), whereas upper secondary school students most often justified the credibility of online texts by referencing the venue or author alongside the evidence (Sub-study II). Identifying the author or publisher of online text and evaluating their expertise in relation to the topic in question reflect fundamental sourcing practices (cf. Bråten et al., 2018b; Perfetti et al., 1999), which are particularly important in the digital world. However, upper secondary school students struggled to evaluate intentions or corroborate online information. Including these evaluation practices in the regular repertoire at the upper secondary school level is therefore important.

Third, upper secondary school students varied in the sophisticated level they had achieved in their reasoning when justifying the credibility of online texts (see Table 5). Almost 10% of the students did not engage in deep reasoning (Sub-study II). In contrast, few students thoroughly justified the credibility of online texts in a regular manner. However, the ability to engage in in-depth reasoning could be expected from students in academic-oriented upper secondary schools (cf. Kiili et al., 2019). Accordingly, deep reasoning demonstrates that the student understands which specific features in the credibility aspects (e.g., author credentials, publication practices, or scientific

intentions) confirm the high credibility of the online text or, reversely, question the credibility of the text.

Finally, upper secondary school students' sourcing varied in the different phases of online inquiry (Sub-study III). Before the intervention, students rarely engaged in sourcing in the first phases when they specified their information need or formulated search queries. For them, these might have been new sourcing practices (see Kiili et al., 2021) that have not often been taught in schools. Further, previous sourcing interventions have not emphasized these practices (see review by Brante & Strømsø, 2018). When upper secondary school students composed their credibility judgments and a written product, some students quite deeply engaged in sourcing practices in their writings, whereas others did not. Evaluating the sources of information and citing sources in essays are probably more familiar sourcing practices for students, although they are not entirely mastered (cf. List et al., 2017; Salmerón et al., 2018a; Strømsø et al., 2013).

7.2 Interventions enhanced students' sourcing skills

The second research question aimed at clarifying the effects of the interventions on students' critical online reading skills. Both interventions of this dissertation enhanced students' sourcing skills in particular. The result is in line with studies at the upper secondary school level (e.g., Braasch et al., 2013; Bråten et al., 2019b; Britt & Aglinskas, 2002) and some studies among young adolescents (e.g., Kingsley et al., 2015; Macedo-Rouet et al., 2013; Zhang & Duke, 2011).

In this dissertation, upper secondary school students participated in a coherent sourcing intervention (Sub-study III), whereas sixth graders' intervention (Sub-study I) was less coherent and mixed, also emphasizing the learning contents. However, the explicit teaching of critical online reading skills with the active practicing of the skills, individually and collaboratively, can promote students' sourcing skills at different educational levels. Sourcing skills are essential in producing the document model (Perfetti et al., 1999) during reading, which helps to compare conflicting information and better understand different issues. In the context of the Internet, sourcing skills are even more evident, as the credibility of online texts can remarkably vary.

However, the intervention among sixth graders did not enhance students' content-related justifications for the credibility of online texts and the use of justifications in their written product (Sub-study I). As these practices are quite demanding, young students might have needed more explicit instruction and time for practicing to learn these skills. However, emphasis on teaching credibility evaluation can be first placed on enhancing younger students' sourcing skills, which are often easier to start with and were also improved during the intervention. Likewise, upper secondary school students (Sub-study III) did not increase their sourcing when specifying their information need, which was only implicitly taught during the intervention. The result highlights the

explicit teaching (cf. Heijltjes et al., 2014; Marin & Helpern, 2011) and practicing of each sourcing skill.

Furthermore, many sixth graders could not write any relevant justifications for the credibility of online texts after the intervention. The result suggests that they might have needed, for example, more time to learn from the model and shared analysis before independently practicing the skills. The intervention also included several online inquiry skills with various aspects to be learned, which might have been overwhelming for young students with highly limited skills. More sequenced instructional materials could enable younger students to concentrate on one basic critical online reading skill, such as evaluation of the author, at the time. In addition, a more coherently designed intervention could decrease young students' cognitive overload and help them in enhancing their skills in a stepwise manner. Further, according to observations, not all sixth graders concentrated on participating, following, and learning from discussions led by teacher and verbal reflections, which could have also been more structured (cf. Applebee et al., 2003; Walraven et al., 2013). Notably, the class teachers of the intervention group could have benefited from longer professional development before the intervention, as implemented for intervention group teachers in Sub-study III.

Among upper secondary school students (Sub-study III), the intervention was differently effective for students with weaker and better skills at the beginning of the study. A highly desirable result was that the students with the weakest skills benefited the most from the intervention. These students probably took advantage of the sourcing examples that teachers introduced in explicit teaching and the collaborative process of online inquiry. The contextualized intervention seems to have helped these students to better understand how sourcing can occur in different, related phases of online inquiry (cf. Kiili et al., 2021). However, for some reasons, the most skillful students did not apply their existing skills in the post-test. Perhaps they would have needed targeted personal feedback during the process and more complex online inquiry tasks to be solved.

Several motivational factors (cf. List & Alexander, 2017; Paul et al., 2017) may also have affected students' performance in pre- and post-tests and learning during the interventions. For example, sixth graders completed two pre-tests and two post-tests and were not probably highly motivated to carefully respond to each test. Upper secondary school students' pre- and post-tests did not affect their course credits, which might have decreased their motivation to respond in detail. Furthermore, some upper secondary school students reported that the investigated topics during the intervention were not highly interesting to them (see Kiili et al., 2022b). As older students were able to form small groups for collaborative work, the constitution of some groups might not have been optimal for learning (cf. review by Wilkinson & Fung, 2002). In addition, observations revealed considerable differences in younger and older adolescents' motivation and concentration during the intervention lessons.

Overall, students' average performance levels could have been higher after both interventions. Despite the role of motivational aspects, more time and

regular practice are probably needed to achieve higher-level learning results in adolescents' critical online reading skills.

7.3 Associations between individual difference factors and students' critical online reading skills varied

The third research question regarded how individual difference factors such as students' gender, basic reading skills, prior topic knowledge, and ISEJ were associated with their critical online reading skills and with the changes in their skills during the intervention (Sub-study III). A recent review by Anmarkrud et al. (2021) showed that these factors have different associations with students' evaluation behavior. In line with this review, most findings in this dissertation were mixed.

Gender. A bit surprisingly, girls and boys at the sixth-grade level performed equally well in justifying the credibility of online texts after the intervention when, for example, their basic reading skills (reading fluency and comprehension) were controlled for (Sub-study I). The results contradict online inquiry studies by Forzani (2018) and Kiili et al. (2018b), where girls and boys differed in their credibility evaluation skills. Instead, Kanniainen et al. (2019) found mixed results among Finnish sixth graders. Particularly in Finland, reading research indicates notable differences between genders in their basic reading skills, favoring girls (e.g., Brozo et al., 2014; Leino et al., 2018; Marôco, 2021).

The result of this dissertation suggests that although basic reading skills form the inevitable basis for critical online reading skills evaluation, the latter may require skills beyond basic reading skills (cf. Coiro, 2011a). A recent study by Sormunen et al. (2021) applied questionnaire data from the same sixth graders as in Sub-study I and found that, interestingly, boys demonstrated more confidence than girls in their searching skills and in the evaluation of search results as well as a more positive attitude toward online inquiry. Thus, practicing critical reading skills in online environments could motivate boys, as compared with girls, they seem to prefer reading online (Liu & Huang, 2008).

Reading fluency. Although the applied test for reading fluency (Holopainen et al., 2004) was similar in all sub-studies, it had different associations with students' critical online reading skills (cf. Anmarkrud et al., 2021). In line with the study by Kanniainen et al. (2019), sixth graders' reading fluency was only approaching statistical significance in having positive association with their credibility evaluation skills after intervention (Sub-study I).

In contrast, similar to the study by Macedo-Rouet et al. (2020), association between students' reading fluency and credibility evaluation skills was found among upper secondary school students (Sub-study II). However, after the intervention, upper secondary school students' reading fluency did not explain

their sourcing skills and hardly explained any changes in their skills during the intervention (Sub-study III). As reading fluency results were mixed throughout the sub-studies, more research is needed to clarify its role in students' critical online reading skills (see also Anmarkrud et al., 2021).

Reading comprehension. The associations between sixth graders' reading comprehension and credibility evaluation skills were also mixed (Sub-study I). The better reading comprehension students had, the better they were at justifying the credibility of online texts (cf. Kanniainen et al., 2019) by referencing the expertise of the source and argumentation in the text. The result replicates previous findings, suggesting that when reading comprehension is measured with open-ended questions, a positive relationship with students' sourcing is often found (e.g., Hahnel et al., 2019; Macedo-Rouet et al., 2013; Salmerón et al., 2020).

Furthermore, sixth graders wrote their justifications for the credibility of online texts, which required writing skills similar to the applied reading comprehension measure. Similarly, a recent study applying the same data as in Sub-study I (Kullberg et al., 2023) showed that sixth graders' performance in the open questions of the reading comprehension measure was associated with their integrative writing in the written product, whereas their performance in the cloze test of the reading comprehension measure was not. Notably, in this dissertation, only data from the open questions of the reading comprehension measure (Kajamies, 2017) were used.

However, sixth graders' reading comprehension did not explain their abilities to justify the credibility of online texts with other source features than the expertise of the source and other aspects of content than argumentation in the text. Evaluating author expertise (e.g., Bråten et al., 2018b) and argumentation, including evidence presented for claims (e.g., Forzani, 2020; Larson et al., 2009; Means & Voss, 1996), can be regarded as deeper-level evaluation criteria than evaluating, for example, the amount of text or the date of publication. Thus, reading comprehension seems to be particularly useful for achieving higher-level evaluation skills.

Prior topic knowledge. Upper secondary school students' prior topic knowledge was only marginally significant in showing a positive association with their abilities to justify the credibility of online texts (Sub-study II). Furthermore, the better prior topic knowledge students had, the more often they engaged in sourcing when formulating search queries after intervention, and vice versa (Sub-study III). In previous studies, students' topic knowledge has played different roles in their sourcing skills depending on how it has been measured (see review by Anmarkrud et al., 2021).

In this dissertation, the prior topic knowledge measure included true/false questions, similar to studies by Kammerer et al. (2016b) and Ulyshen et al. (2015), which also did not find associations with upper secondary school students' sourcing skills. The applied measure could be relatively narrow to comprehensively cover students' prior topic knowledge. Accordingly, McCarthy and McNamara (2021) argued that prior knowledge measures should cover

amount, accuracy, specificity, and coherence. For example, specificity refers to the degree to which prior topic knowledge is related to the information in the texts to be read (McGarthy & McNamara, 2021), which was not checked beforehand when developing the measure for topic knowledge in Sub-studies II and III. Based on the dimensions mentioned above, the prior topic knowledge measure can be improved in future studies.

Internet-specific epistemic justifications (ISEJ). The results of Sub-study II confirmed some previous findings of ISEJ among older students (Kammerer et al., 2013, 2020). Students who believed that they evaluated the authority or compared documents when they read online were better at justifying the overall credibility of online texts, and vice versa. This result is desirable because the above-mentioned epistemic beliefs present higher-level evaluation practices than personal justification as personal knowledge can sometimes include false beliefs or biased information (Greene et al., 2019).

The result also suggests that although many students overestimate their critical reading skills (Paul et al., 2017), epistemic awareness of their own useful evaluation practices may enhance credibility evaluation of online texts (cf. Bråten et al., 2022). Sub-study II investigated how students' epistemic beliefs were associated with their overall evaluation performance. Future studies could investigate how each ISEJ dimension (e.g., authority, corroboration) is associated with students' actual use of corresponding evaluation criteria.

In sum, the mixed results indicate that research should further investigate how individual difference factors affect adolescents' critical online reading skills and the learning of those skills (see also Anmarkrud et al., 2021). However, the results of this dissertation established that students' basic reading skills are not without matter in their critical online reading skills (see also Kanniainen et al., 2019) and that students' epistemic beliefs about justifications for knowing play a significant role in their skills.

7.4 Topic and students' text selections explained their critical online reading skills

Although students' prior topic knowledge was hardly associated with their critical online reading skills, the topic of the online inquiry task was significant in all sub-studies. Previous studies have shown that the topic of the reading materials seems to affect students' evaluation behavior (e.g., Bråten et al., 2018b). Further, the examined topic can modify how individual difference factors are associated with students' critical reading skills (see Anmarkrud et al., 2021).

In this dissertation, upper secondary school students who explored the fats topic performed better in justifying the credibility of the online texts than those who investigated the vaccination topic (Sub-study II). Similarly, students' prior topic knowledge was higher and selecting the most useful online texts was easier on the fats topic than on the vaccination topic. However, the associations

between students' ISEJ and their evaluation performance did not differ according to the topic.

In contrast, upper secondary school students' sourcing in credibility judgments was more common in the vaccination topic than in the fats topic (Sub-study III). In fact, students who explored the vaccination topic engaged more often in sourcing in all phases of online inquiry than those who examined the fats topic. Thus, this difference also explained why sourcing performance of some more advanced students worsened during intervention. Previous studies (e.g., Bråten et al., 2018b; Lucassen et al., 2013; Lucassen & Schraagen, 2011; Stadtler & Bromme, 2014) have shown that sourcing may be crucial when readers deal with topics that are less familiar to them, and on which they have less prior knowledge, and this might explain some topic differences in this dissertation.

Interestingly, however, before the first online inquiry task, upper secondary school students exploring the vaccination topic self-evaluated themselves as having significantly more knowledge about the topic than students exploring the fats topic. Furthermore, after completing the task, although students reported that locating texts was significantly easier in the fats topic than in the vaccinations topic, no perceived differences were observed between topics in terms of difficulty of credibility evaluations. As students' self-evaluations mostly contradicted the results obtained in this dissertation, self-assessments seem to not be valid in measuring students' prior topic knowledge or critical online reading skills.

In addition, the topic order was influential in both intervention studies (Sub-studies I and III). For example, after the intervention, sixth graders who investigated the computer gaming topic more often justified the credibility of the online texts by referencing argumentation in the text than sixth graders who examined the reading on screen topic (Sub-study I). The computer gaming topic might have been more interesting for students to investigate and evaluate in detail than the reading on screen topic, although students' self-reported prior topic knowledge did not differ between topics.

Further, upper secondary school students' text selections were positively associated with their abilities to justify the credibility of online texts (Sub-study II). That is, the more useful online texts students selected, the better they were in evaluating various credibility aspects and engaging in deep reasoning when justifying the credibility of those texts, and vice versa. It seems obvious that selection and evaluation skills are related to each other, as online readers need to evaluate credibility when selecting online texts from search engine results pages (SERPs), and it continues after opening the Web pages to read the texts more closely (e.g., Gerjets et al., 2011; Rieh, 2002).

8 GENERAL DISCUSSION

This dissertation contributed to a significant societal phenomenon since false information spreading online challenges our abilities to make decisions based on credible – scientific or expert-delivered – information (cf. Ecker et al., 2022). Despite recognizing the problem, it is extremely difficult to prevent incorrect information on the Internet – regardless of whether it has been shared mistakenly or intentionally. Therefore, this dissertation highlighted critical online reading skills, including skills to recognize and avoid false online information and abilities to locate and take advantage of more credible information. By developing instructional materials and methods for schools, adolescents' critical online reading skills can be improved, and thus, the shortcomings of incorrect online information diminished.

This dissertation furthered our knowledge of critical online reading in at least three areas. First, it increased our understanding of adolescents' critical online reading skills, particularly their abilities to justify the credibility of online texts and engage in sourcing during different phases of online inquiry. Second, the dissertation provided teachers and educators with comprehensive and accessible instructional materials and methods to enhance critical online reading skills among students of different ages. Third, it shed light on individual differences and topic-related factors that may affect adolescents' critical online reading skills and their learning of those skills.

This dissertation also made significant methodological and pedagogical contributions. The extensive pre- and post-test data of students from two different educational levels, comprehensive interventions included in the regular school curricula, and detailed analyses of students' responses from versatile methodological approaches allowed unique methodological and pedagogical contributions and implications that I will discuss next. Finally, I evaluate the dissertation and make suggestions for future research.

8.1 Methodological contributions and implications

This dissertation makes several significant methodological contributions that may advance the research on critical online reading. These contributions include the methods developed for data analysis, validation of the ISEJ inventory (Bråten et al., 2019a) among Finnish upper secondary school students, and customized search engines embedded in Web-based environments of online inquiry tasks.

First, after counting all relevant, mainly superficial sixth graders' justifications for the credibility of online texts in Sub-study I, a more targeted scoring rubric for upper secondary school students' justifications was designed (Sub-study II). The rubric takes into account the most crucial credibility aspects in online texts and the depth of students' reasoning. Based on previous studies (e.g., Bråten et al., 2018b; Kohnen & Mertens, 2019; Sinatra & Lombardi, 2020), the scoring rubric highlights the most relevant credibility aspects in online texts (author, venue, intentions, evidence, and corroboration), which upper secondary school students should be able to pay attention to in their justifications.

Moreover, the scoring rubric distinguishes how superficially or deeply students engage in reasoning (cf. Coiro et al., 2015; Kiili et al., 2019). Although scoring students' written responses according to the multi-stage rubric is challenging, it accurately reveals the possible variation in upper secondary school students' critical online reading skills. Therefore, future studies among older adolescents should not only examine how students acknowledge the most relevant credibility aspects in online texts but also investigate the level of their reasoning. These criteria may unify the analysis protocols and, thus, make results more comparable across different studies.

Second, a unique methodological contribution of this dissertation was the use of the RCI (Jacobson & Truax, 1991) in the analysis of the efficacy of the intervention in Sub-study III. The RCI allows the development of more efficient teaching methods and interventions by discovering, in detailed ways, what kinds of students benefit from the intervention. RCI analysis derives from studies investigating the clinical significance of therapy (see Jacobson & Truax, 1991) and has been rarely used in reading interventions (cf. Aro et al., 2018). Most intervention studies focusing on reading examine intervention effects only at the group level without elaborating on those who benefit from the intervention (see review by Brante & Strømsø, 2018). Nevertheless, examining for whom the intervention works is crucial. With this information, instructional methods and practices can be further developed to better serve all learners.

Further, after students are categorized into RCI classes, the associations between their RCI classes and different individual differences or topic-related factors can be examined. Accordingly, the analysis shows whether students with specific characteristics (e.g., lower or higher levels of reading fluency and prior topic knowledge) are more likely to improve, not change, or worsen their critical online reading performance during the intervention. Future studies, using RCI (Jacobson & Truax, 1991), may more widely explore which motivational,

cognitive, and affective factors are related to changes in students' critical online reading skills during the intervention and take them into account when developing more efficient teaching methods.

When taking advantage of RCI analysis, students' skills must also be assessed before the intervention, which has not been consistently performed in previous intervention studies (see Brante & Strømsø, 2018). In this dissertation, RCI analysis revealed that students with the weakest skills benefited the most from the intervention. It also exposed the challenges in motivating the more advanced students to maintain or improve their critical online reading skills. These results remarkably expand the understanding of how selected instructional methods during intervention enhanced upper secondary school students' critical online reading skills compared with group-level analyses showing only a positive effect in three of four outcome variables. Future intervention studies should further this research and utilize both group-level investigations and more individual approaches, such as RCI analysis (Jacobson & Truax, 1991), to increase the applicability of the results.

Third, this dissertation contributed by validating the ISEJ inventory (Bråten et al., 2019a) among Finnish upper secondary school students across two health topics (Sub-study II). Most previous studies using ISEJ have been conducted among university students (Binali et al., 2021; Bråten et al., 2019b; Kammerer et al., 2021), whereas in this dissertation, the inventory was successfully applied to measure younger students' epistemic beliefs (see also Cheng et al., 2021). Accordingly, a three-dimensional structure of ISEJ, including justification by authority, justification by multiple sources, and personal justification, was also found among Finnish upper secondary school students (see also Kiili et al., 2022a).

As the fourth significant methodological contribution, customized search engines incorporated in the Web-based environments of the online inquiry tasks enabled nearly authentic Web search experiences for adolescents but limited their text selections (Sub-studies I–III). Moreover, the customized search engines collected log data for researchers, allowing the investigation of all students' attempts to formulate search queries. Furthermore, students' responses in the following task phases were easier to compare when their text selections, evaluations, and written products were based on a limited number of different online texts.

As the aim of this dissertation was to investigate adolescents' critical online reading skills, the use of authentic online texts provided genuine text selections and evaluations. However, a limited number of online texts resulted in SERPs, where for some topics, the most useful texts were more often provided as the first ones in the list compared with other topics. In addition, some upper secondary school students navigated outside the search engines by following the hyperlinks in pages. Despite these challenges, the use of customized search engines is a promising practice to be further developed.

8.2 Pedagogical contributions and implications

The main pedagogical contribution of this dissertation is that teaching critical online reading skills, including evaluation and sourcing practices, can be coherently integrated as a part of online inquiry (cf. Leu et al., 2019; see also Kiili et al., 2022b). At the sixth-grade level, the process was less coherent (Sub-study I), whereas at the upper secondary school level, the intervention was the first to systematically teach sourcing in different phases of online inquiry (Sub-study III). In previous intervention studies (see review by Brante & Strømsø, 2018), sourcing has been highlighted when students evaluate the credibility of texts or use source information to synthesize information in written products. Limited studies have emphasized the role of sourcing in defining information need or formulating search queries (see also Kiili et al., 2021). In this dissertation, the authentic integrated process of online inquiry during intervention helped older adolescents to understand how the different phases and critical online reading practices are intertwined.

In this dissertation, different instructional methods were integrated into a practical unit in ways where repetitive structures made the lessons more predictable for students. At the sixth-grade level, the explicit teaching of each online inquiry began with video modeling (e.g., Choi & Johnson, 2005), followed by students practicing the modeled skill (Sub-study I), whereas teachers at the upper secondary school level first gave a short introduction to each online inquiry skill, followed by students practicing the skill in small groups (Sub-study III). At both educational levels, modeling or explicit teaching covered the skills needed when searching for information, evaluating the credibility of information, and synthesizing information. Despite the integration of several instructional methods in the interventions of this dissertation, modeling or explicit teaching seem to be an invaluable part of the efficient teaching of critical online reading skills (e.g., Heijltjes et al., 2014; Marin & Helpern, 2011).

Another influential pedagogical contribution was the successful use of a shared working document by upper secondary school students when working in small groups (Sub-study III; see also Kiili et al., 2022b). Compared with separate worksheets in the sixth graders' intervention (Sub-study I), a single digital document including all task prompts and questions not only guided students' work throughout the lessons but also documented their online inquiry process and made it accessible and visible to the teacher. Note that the document covering the entire process throughout the lessons was suitable for older adolescents but could have caused excess cognitive load for younger adolescents, such as sixth graders.

A joint document requires students to collaboratively share their thoughts, reflect on their learning, and decide how to proceed with the sub-tasks—that is, metacognitively plan and guide their process according to the prompts in the working document (cf. Stadtler & Bromme, 2007). The document, including all phases of online inquiry, emphasizes the role of the learning process instead of

the result only. An interview study among adolescents by Paul et al. (2017) suggested that when sourcing is actively and concretely prompted and appreciated by the teachers, students more often engage in it and learn to regularly accomplish it.

As a specific pedagogical contribution, the developed scoring rubric for upper secondary school students provides teachers with a tool to analyze their students' skills and to help them in comparing superficial and more advanced evaluations. As many students lacked the skills to regularly justify the credibility of online texts in a thorough manner, there is a need to teach how deep evaluation occurs during online reading. Accordingly, in-depth reasoning in credibility judgments (cf. Kiili et al., 2019) requires the careful examination of source features in texts and understanding what counts for convincing, credible evidence and the importance of corroborating found information with credible documents. The scoring rubric can also be simplified using, for example, three levels of reasoning instead of the four levels applied in Sub-study II.

Above-mentioned pedagogical contributions and implications serve not only teaching critical online reading skills for adolescents but also when there is a need to enhance older students' skills. For example, many pre-service teachers seem to struggle with justifying the credibility of online texts (Kulju et al., under review). Thus, almost a similar coherent process of online inquiry and a joint working document with a developed scoring rubric to assess the skills could be applied for improving their skills. Furthermore, students' critical online reading skills should not only be developed in schools as a part of language arts but as an important part of almost all school subjects (cf. National Core Curriculum for Basic Education, 2014; National Core Curriculum for General Upper Secondary Education, 2019). By selecting different issues and topics for students to investigate, the instructional materials and methods of the interventions in this dissertation could be applied and further developed in many school subjects. For example, in Sub-study I, sixth-graders' other project (8 x 45 min lessons) considered energy and was embedded in school subject Environmental studies.

Although the interventions in this dissertation were shown to be feasible, there is always room for improvement. Different scientific methods or strategies (e.g., action research, design research) could help in planning and further improving interventions and their instructional methods. In future studies, more attention should also be paid, for example, to adolescents' epistemic beliefs, differentiating instruction, motivating students, and personal feedback.

First, even though upper secondary school students' ISEJ (Bråten et al., 2019a) were measured before the intervention, they were not leveraged during the intervention. The found positive associations between students' evaluation performance and their epistemic beliefs about the justification by authority and multiple sources suggest that, at least, older adolescents are aware of their own evaluation practices. This awareness could be a starting point for intervention when students critically consider and reflect on their own epistemic practices to justify the credibility of online texts. A recent study (Bråten et al., 2022) showed that a refutation text intervention changed university students' epistemic beliefs,

which was also transferred to various stages of their multiple document task. Thus, future interventions could aim to affect students' critical online reading skills, their Internet-specific epistemic beliefs, and the reciprocal relationship of beliefs and skills.

Second, as this dissertation found considerable variation in adolescents' critical online reading skills, teaching should be differentiated at every educational level. Critical online reading skills are shown to be challenging for most young adolescents in particular (e.g., Coiro et al., 2015; Forzani, 2018; Kiili et al., 2018b); thus, more time could be devoted to teaching and practicing basic skills with concrete examples. At the same time, more advanced young adolescents could benefit from deepening their understanding of the different credibility aspects and their relationships. Similarly, more advanced older adolescents could be given more demanding controversial topics to investigate, be tasked to practice corroborating the information, or respond to questions requiring a deeper level of thinking and reasoning (e.g., how source features of online texts may affect interpreting the content). Regarding the role of individual differences in adolescents' critical online reading skills, more research is needed to take these into account when differentiating teaching (cf. Anmarkrud et al., 2021).

Third, as in the critical online reading interventions in general (see Brante & Strømsø, 2018), more attention should be paid to motivating students' learning and their performance in tasks measuring their skills. In this dissertation, many students did not improve their critical online reading performance during the interventions. Previous studies have shown that students can be overconfident about their sourcing skills (e.g., Paul et al., 2017), which may affect their motivation to learn and apply new skills. Thus, this affects how their critical online reading skills develop during teaching.

Moreover, behavioral engagement plays a critical role in developing critical online reading skills (cf. Bråten et al., 2018a; List & Alexander, 2017). When students are motivated, they invest more time and effort into their performance and learning, which leads to better learning results. Accordingly, the recent study by Bråten et al. (2021) showed that university students' writing time and the length of students' written responses had distinctive, unique effects on their reading comprehension performance and mediated the effects of cognitive (e.g., prior knowledge) and motivational individual differences on comprehension performance.

Finally, the differences in adolescents' critical online reading skills accentuate the role of personal feedback during teaching (Hattie & Timperley, 2007; Van der Kleij et al., 2015), which is adjusted to each student's competence level. During the interventions of this dissertation, specific and regular feedback mechanisms were not applied; thus, students' personal feedback depended on the teacher's time resources and intuitive abilities. Effective feedback requires that students understand the learning goals and what constitutes the advanced skills against which their performance can be compared (Nicol & Macfarlane-Dick, 2006). By modeling and instructing, desired skills can be demonstrated

with concrete examples, but personal feedback is also needed while independently practicing those skills (cf. Nicol & Macfarlane-Dick, 2006).

The meta-analysis by Van der Kleij et al. (2015) showed that the method of providing feedback to students is also essential. Elaborated feedback, such as providing hints, additional information, extra study material, or an explanation of the correct answer (Shute, 2008), was more effective than feedback regarding the correctness of the answer or the correct answer provided. Moreover, elaborated feedback was particularly effective for higher-order learning outcomes (Van der Kleij, 2015). However, because of large class sizes and teachers' limited time resources, alternative ways of providing feedback, such as digital or peer feedback, are worth developing.

Regular personal feedback enables students' more detailed and profound level of self-assessment and reflection (Nicol & Macfarlane-Dick, 2006). Pedagogical designs could use specific reflection prompts that are carefully aligned with the learning objectives and explicitly related to the skills that students need to learn. Likewise, ensuring sufficient time for self-assessment and reflection is vital. Alongside personal feedback, these could help students at different competence levels to provide a more realistic view of their learning needs and develop their critical online reading skills.

8.3 Evaluation and future directions for research

This section evaluates the dissertation by discussing the challenges faced when designing and implementing the interventions, measures, and analyses of this dissertation. The specific limitations of this dissertation are also addressed, followed by the directions for future research.

One of the challenges in assessing and teaching students' critical online reading skills is to find suitable topics for online inquiry tasks before, during, and after the intervention. The selection of topics is crucial, as different topics elicit students' critical reading skills somewhat differently (cf. Bråten et al., 2018b), which was confirmed in this dissertation. Several requirements need to be met when selecting the topics. First, to maintain students' engagement through the task, the applied topics must be interesting for adolescents (cf. Chinn et al., 2021). According to students' and teachers' feedback (see Kiili et al., 2022b), this was not sufficiently ensured during interventions in this dissertation. Second, as controversies in texts seem to enhance students' sourcing (e.g., Kammerer et al., 2016a), controversial topics need to be selected, thus restricting topic selections. Particularly for younger students, the controversy in texts should be clear and easily interpreted. Third, when authentic online texts are applied, ensuring that a sufficient variety of texts on the topic is available on the Internet is essential. Finally, even more difficult is to find topics that can be used as alternatives in pre- and post-tests. As discovering similarly working topics for pre- and post-tests is generally difficult, statistical analyses should be used to control for the role of the topic.

Second, this dissertation addresses both the advantages and challenges of using researcher-designed texts or applying authentic online texts in assessing students' critical reading skills. For example, using researcher-designed online texts in Sub-study I allowed the manipulation of the text variation according to, for example, author, venue, and position on the topic. It also helped to score and compare sixth graders' responses when they wrote justifications for credibility or synthesis based on the same three texts. However, the online texts in sixth graders' online inquiry task could have included clearer controversies (e.g., Kammerer et al., 2016b), as highly non-credible texts were not designed. In addition, some students reasonably questioned the credibility of, for example, Web news designed by researchers—such as the venue of which they had never heard. Note that a closed search engine was incorporated into the Neurone environment (González-Ibañez et al., 2017; Sormunen et al., 2018), which enabled a quite authentic information search protocol, even though the same texts were given to students after the search process. Furthermore, the online texts were designed to, for example, visually meet the characteristics of authentic online texts. These aspects might have enhanced sixth graders' feelings of authentic online reading.

Among upper secondary school students, only authentic online texts were used in the pre- and post-tests and during the intervention. When the aim is to study or teach critical online reading skills, original online texts are highly preferable, as they allow genuine Internet reading experiences for students. Although the online texts were preselected for the pre- and post-tests and the Google custom search engine limited students' selections, some challenges emerged. Sub-study II revealed that many students followed the hyperlinks in the preselected pages. However, rubrics for upper secondary school students' justifications regarding the credibility of online texts were adjusted to enable equal scoring across the three online texts the students had selected from inside and/or outside the Google custom search engine. Although challenges may arise when using authentic online texts, future studies should develop designs where authentic online reading experiences can be realized.

The online texts that students can select, analyze, and synthesize are also significant during the interventions. For sixth graders, the texts were researcher-designed when the explicit teaching of online inquiry skills occurred. The research-designed texts enabled adjustments to the texts based on students' age levels and the aims of the lessons. Furthermore, sixth graders investigated authentic online texts when they practiced modeled skills in restricted and open online environments during the projects. However, upper secondary school students searched for online texts on the open Web during the entire intervention. The intervention group teachers reported that students had difficulties understanding the concept of "stakeholder" in the task assignment, leading them to select online texts without a clear controversy (see Kiili et al., 2022b). As adolescents have more difficulties in questioning the credibility of online texts than in confirming their credibility (Kiili et al., 2023), school tasks should also include reading less credible online texts that need to be questioned.

Third, the present study included extensive data on students' responses to open questions during different phases of online inquiry. Using written responses as a basis for analysis has advantages and disadvantages. It may reveal, for example, evaluation criteria that adolescents spontaneously use when reading online (cf. Walraven et al., 2009), Further, students' written responses may include aspects that researchers did not assume beforehand and, thus, allow more data-driven analysis. However, open questions are challenging for students to answer when compared with, for example, multiple-choice or ranking items. In addition, students might not be motivated to write detailed responses. Previous research has also shown that adolescents are more capable of expressing their sourcing skills when they are, for example, interviewed (e.g., Macedo-Rouet et al., 2019). Writing long responses might be challenging and time-consuming, particularly for younger students.

From the researcher's point of view, analysis of hundreds of students' written responses is time-consuming. Furthermore, when analyzing students' responses, their interpretations may vary; thus, sufficient inter-rater reliability might be challenging to achieve. Future studies could develop online inquiry tasks mainly based on multiple-choice, rating, or ranking items complemented with open questions. This dissertation offers information about students' critical online reading skills and criteria, which can be used to develop items for those tasks.

In addition to the above-introduced methodological challenges, this dissertation has some specific limitations. First, the measured individual differences were mostly cognitive and did not include, for example, motivational factors (cf. List & Alexander, 2017), such as task values, achievement goals, or self-concept of ability (see also the review by Anmarkrud et al., 2021), or other sociodemographic factors than students' gender. As the associations with adolescents' critical online reading skills were mostly mixed, some other individual factors could have explained part of the results regarding students' performance in online inquiry tasks or their learning during the interventions.

Second, the class teachers of the intervention group of sixth-graders did not receive specific professional development for teaching critical online reading skills before they implemented the lessons of the intervention. As these skills have not necessarily been emphasized upon in their teacher education, they would have needed more support in teaching these skills. Further, professional development at the upper secondary school level could have been more comprehensive for intervention group teachers to reach a profound understanding of sourcing in online reading (cf. Bråten et al., 2019b).

Third, the intervention group teachers at both educational levels could also have been more engaged in planning the instructional materials and methods of the interventions. Then, they might have been more motivated to implement the intervention, for example. However, the time schedules and resources of the projects and teachers' own basic work restricted their involvement.

Fourth, even though the fidelity of the implemented interventions was assured in several ways (cf. McKenna et al., 2014), a score based on fidelity could have been used as a moderating variable in the statistical analyses. This analysis would have revealed if differences in teachers' implementation of the intervention moderated students' learning during the intervention.

Fifth, information about teachers, such as their qualifications, teaching experience, and motives for participating in the study, was not collected. However, students' class was used as a clustering variable in the statistical analysis, which takes into account the differences among the teachers. But based on this analysis, the origin of the variance (is it due to teacher or other factors in the class) could not be interpreted with its meaning for students' learning.

Sixth, because the interventions already included many lessons embedded in regular schoolwork, implementing a delayed post-test to investigate how permanent were the achieved learning gains was not possible (cf. Bråten et al., 2019b). Future studies may investigate the stability of the changes at the individual level by taking advantage of the RCI analysis (Jacobson & Truax, 1991).

SUMMARY IN FINNISH

Tutkimuksen tavoitteet

Tämän väitöstutkimuksen taustalla on merkittävä eriarvoisuuteen vaikuttava yhteiskunnallinen ongelma eli Internetissä leviävän epäluotettavan informaation lisääntyminen, mikä vaikeuttaa luotettavan tiedon tunnistamista ja siihen pohjautuvaa päätöksentekoa. Näin ollen tutkimuksen tavoitteena oli selvittää, voidaanko tutkivan nettilukemisen interventioilla edistää nuorten kriittisen nettilukemisen taitoja, sekä kehittää menetelmiä taitojen opettamisen tueksi. Tutkimuksessa kriittisellä nettilukemisella tarkoitetaan taitoja arvioida tekstien luotettavuutta (ks. Barzilai ym., 2020) sekä taitoja tunnistaa ja hyödyntää lähteiden piirteitä kuten kirjoittaja, julkaisija ja motiivit (esim. Bråten ym., 2018b; Perfetti ym., 1999) tutkivan nettilukemisen eri vaiheiden aikana (Leu ym., 2019). Tutkimuksen tavoitteena oli myös saada tarkempaa tietoa kuudesluokkalaisten ja lukiolaisten kriittisen nettilukemisen taidoista sekä kehittää menetelmiä taitojen analysointiin. Lisäksi väitöstutkimuksessa selvitettiin, miten yksilölliset tekijät (sukupuoli, peruslukutaidot, aiempi tieto aiheesta ja episteemiset uskomukset) sekä tehtäväaihe ja tekstivalinnat olivat yhteydessä nuorten kriittisen nettilukemisen taitoihin.

Tutkimuksen toteutus

Väitöstutkimuksen aineisto on kerätty Suomen Akatemia rahoittamissa tutkimusprojekteissa kuudennella luokalla (iFuCo-hanke 2016–2018; päätösnumero 294197) ja lukiossa (Aroni-hanke 2015–2019; päätösnumero 285817). Molemmilla kouluasteilla toteutettiin interventio, jossa nuorille opetettiin tutkivan nettilukemisen taitoja.

Luokanopettajat toteuttivat interventioon kuuluvat oppitunnit (21 × 45 min) interventioryhmän kuudesluokkalaille (N = 190), kun taas kontrolliryhmä (N = 152) osallistui normaaliin kouluopetukseen. Ensimmäisessä opetuskokonaisuudessa opetettiin tutkivan nettilukemisen taitoja. Aluksi oppilaille mallinnettiin yhtä tutkivan nettilukemisen osataitoa (tiedonhaku, nettitekstien luotettavuuden arviointi tai synteesi) videon avulla, minkä jälkeen oppilaat harjoittelivat taitoa itsenäisesti, taidosta keskusteltiin yhdessä ja lopuksi reflektoitiin opittua. Kunkin osataidon opettamisen jälkeen (yhteensä 9 × 45 min) oppilaat pääsivät harjoittelemaan tutkivan nettilukemisen taitoja kahdessa projektissa, joista ensimmäinen tapahtui suljetussa ja toinen avoimessa nettiympäristössä.

Lukiolaisten interventio (4 × 75 min) oli osa suomen kielen ja kirjallisuuden opettajan toteuttamaa Tekstit ja vaikuttaminen -kurssia ja kontrolliryhmän opiskelijat osallistuivat vastaavalle tavalliselle kurssille. Ensimmäisellä interventionoppitunnilla opiskelijat valitsivat heitä kiinnostavan terveysaiheen sekä pienryhmän (2–4 opiskelijaa), jossa työskenneltiin koko intervention ajan. Kolmella ensimmäisellä oppitunnilla opettaja esitti lyhyet tietoiskut tiedonhausta, nettitekstien luotettavuuden arvioinnissa sekä synteesistä. Kunkin tietoiskun jälkeen opiskelijat työskentelivät pienryhmissä vastaamalla yhteisen työskentely-

dokumentin kysymyksiin ja tehtäviin. Viimeisen oppitunnin seminaarissa opiskelijat jakoivat toisilleen tutkivan nettilukemisen prosessinsa tuloksia ja oppimaansa.

Alku- ja lopputestinä sekä kuudesluokkalaiset että lukiolaiset tekivät tutkivan nettilukemisen tehtävän. Tehtävänannossa heitä pyydettiin tutkimaan ristiriitaista aihetta suljetussa nettiympäristössä. Kuudesluokkalaisten aiheina olivat tietokonepelaaminen ja ruudulta lukeminen ja lukiolaisten aiheina olivat rokottaminen ja ravintorasvat. Tehtävässä nuoret etsivät ja valitsivat nettitekstejä täsmähakukoneen avulla, arvioivat nettitekstien luotettavuutta ja kirjoittivat synteesin nettitekstien avulla. Lisäksi nuoret tekivät peruslukutaitoja mittaavan testin (Holopainen ym., 2004; Kajamies, 2017) ja lukiolaisilta kartoitettiin myös heidän Internet-spesifit episteemiset uskomuksensa (Bråten ym., 2019a) sekä aiempi tieto tehtäväaiheista.

Väitöstutkimuksen laadullisen aineiston muodostivat kuudesluokkalaisten ja lukiolaisten vastaukset tutkivan nettilukemisen tehtävän eri vaiheissa. Nettitekstien luotettavuuden perusteluista tarkasteltiin, mitä lähteiden piirteiden (esim. Bråten ym., 2018b; Britt & Aglinskas, 2002) tai nettitekstien sisällön (esim. Braasch ym., 2013; Britt ym., 2014; Kohnen & Mertens, 2019) arviointikriteerejä nuoret käyttivät. Yksittäinen lause saattoi sisältää yhden tai useamman arviointikriteerin. Kuudesluokkalaisten vastauksista tunnistettiin kaikki relevantit nettitekstien luotettavuuden arviointikriteerit. Lisäksi arviointikriteerit luokiteltiin neljään luokkaan: lähteen asiantuntijuus, muut lähteen piirteet, tekstin argumentaatio ja muut sisällölliset piirteet. Sen sijaan lukiolaisten vastauksista huomioitiin vain tärkeimpien arviointikriteerien (kirjoittajan, julkaisijan, motiivien ja evidenssin arviointi sekä korroboraatio) käyttö ja arviointikriteerin käytön syvällisyys. Näin ollen lukiolaisten arviointipistemäärä huomioi sekä heidän perusteluidensa monipuolisuuden että syvällisyyden.

Lisäksi kuudesluokkalaisten kirjoitelmista (synteesi) tarkasteltiin, hyödynsivätkö he luotettavuuden perusteluissa käyttämiään arviointikriteerejä myös kirjoitelmassaan. Lukiolaisten tiedontarpeen määrittelystä ja hakulausekkeista analysoitiin, miten he hyödynsivät niissä lähteiden piirteitä kuten nettitekstien kirjoittajaa ja julkaisijaa. Myös lukiolaisten tekemät nettitekstivalinnat pisteytettiin valintojen hyödyllisyyden (relevanssi ja luotettavuus) mukaisesti. Lisäksi lukiolaisten kirjoitelmista analysoitiin, miten he esimerkiksi viittasivat lähteisiin ja muodostivat lähde–lähde tai lähde–sisältö-linkkejä (vrt. Perfetti ym., 1999).

Kvantifioitu aineisto analysoitiin monipuolisten tilastollisten menetelmien avulla. Esimerkiksi interventioiden tehokkuutta tutkittaessa regressioanalyyseissa kontrolloitiin useita taustatekijöitä (alkutestin pistemäärä, sukupuoli, peruslukutaidot, tehtäväaihe, aiempi tieto aiheesta), jolloin pystyttiin myös tarkastelemaan erikseen näiden tekijöiden yhteyttä nuorten kriittisen nettilukemisen taitoihin. Lisäksi Reliable change index -analyysilla (RCI; Jacobson & Truax, 1991) pystyttiin selvittämään, millaiset opiskelijat hyötyivät interventiosta.

Tulokset ja johtopäätökset

Väitöstutkimuksessa löydettiin huomattavia eroja niin kuudesluokkalaisten kuin lukiolaistenkin kriittisen nettilukemisen taidoissa. Eroja nuorten välillä huomattiin esimerkiksi: 1) arviointikriteerien käytössä, 2) perusteluiden syvällisyydessä ja 3) lähteiden piirteiden hyödyntämisessä tutkivan nettilukemisen prosessin eri vaiheissa. Koska selkeitä taitoeroja löydettiin jo kuudesluokkalaisilla, olisi tärkeää selvittää tarkemmin, miten ja missä vaiheessa erot kehittyvät sekä esimerkiksi mitkä kotiympäristön tekijät vaikuttavat taitojen kehittymiseen. Lisäksi kriittisen nettilukemisen perustaitoja tulisi opettaa jo alakoulussa riittävän varhaisessa vaiheessa. Tulosten perusteella lukiolaisten opetus tulisi kohdentua informaation varmentamiseen muiden tekstien avulla (esim. Kohnen & Mertens, 2019) sekä syvälliseen arviointikriteerien käyttöön (esim. Kiili ym., 2019).

Väitöstutkimus osoitti, että taitojen mallintamisella ja eksplisiittisellä opettamisella sekä yksilöllisellä ja yhteisöllisellä harjoittelulla tutkivan nettilukemisen prosessin aikana voidaan parantaa nuorten kriittisen nettilukemisen taitoja, erityisesti lähteiden piirteiden (esim. kirjoittaja, julkaisija) huomioimista, arvioimista ja hyödyntämistä. Tulos on tärkeä, sillä esimerkiksi kirjoittajan asiantuntijuuden arvioiminen on olennainen osa kriittisen nettilukemisen taitoja (vrt. Bråten ym., 2018b; Perfetti ym., 1999) ja nettitekstien sisällön arviointia helpompi taitojen opettamisen lähtökohta.

Sen sijaan interventio ei parantanut kuudesluokkalaisten taitoja perustella nettitekstien luotettavuutta sisällöllisillä tekijöillä (esim. argumentoinnin tasapuolisuus, evidenssin laatu, kirjoitustyyli) tai hyödyntää perusteluja kirjoitelmissaan. Lukiolaisilla interventio ei puolestaan lisännyt lähteiden piirteiden hyödyntämistä tiedontarpeen määrittelyssä. Nämä tulokset olivat osin odotettuja, sillä esimerkiksi nettitekstien sisällön luotettavuuden arviointi on haastava taito oppilaille, varsinkin jos aiempaa tietoa aiheesta ei ole paljon (esim. Bromme & Goldman, 2014). Lisäksi lukiolaisten interventiossa tiedontarpeen määrittelyn tärkeys oli mukana vain implisiittisesti eli tulos korostaa taitojen eksplisiittisen opettamisen tärkeyttä (vrt. Heijltjes ym., 2014; Marin & Helpern, 2011).

Toisaalta väitöstutkimuksessa havaittiin, että intervention jälkeen niin kuudesluokkalaisten kuin lukiolaistenkin kriittisen nettilukemisen taitojen keskimääräinen taso olisi voinut olla korkeampi. Tulosta selittävät luultavasti myös monet motivaatiotekijät (vrt. List & Alexander, 2017; Paul ym., 2017), sillä observoinnit ja opettajien näkemykset (ks. Kiili et ym., 2022) viittasivat siihen, että oppituntien aikaisessa sitoutumisessa ja keskittymisessä oppimiseen oli isoja eroja nuorten välillä. Kaikki nuoret eivät myöskään luultavasti motivoituneet vastaamaan lopputestiin niin hyvin kuin olisivat osanneet, sillä esimerkiksi kuudesluokkalaisilla samanlainen testi toistui yhteensä neljä kertaa, ja testissä suoriutuminen ei vaikuttanut arvosanoihin kummallakaan kouluasteella. Motivaatiotekijöistä huolimatta laajempien ja korkeatasoisempien oppimistulosten saavuttamiseksi tarvitaan luultavasti toistuvaa ja pitempiaikaista kriittisen nettilukemisen taitojen opetusta.

Väitöstutkimuksessa nuorten kriittisen nettilukemisen taitoihin olivat yhteydessä alkutestitulokset, tehtäväaihe, tekstivalinnat sekä episteemisistä usko-

muksista nettitekstin asiantuntijuuden arvioiminen ja informaation varmentaminen muiden tekstien avulla (vrt. Kammerer ym., 2013, 2020). Peruslukutaidoilla (lukusujuvuus, luetun ymmärtäminen) oli merkitystä nuorten kriittisen lukemisen taidoissa (vrt. Coiro, 2011a; Kanniainen ym., 2019, 2022; Kiili ym., 2018a), mutta yhteydet eivät olleet yksiselitteisiä. Sen sijaan tässä väitöstutkimuksessa sukupuoli tai aiempi tieto aiheista eivät selittäneet nuorten kriittisen nettilukemisen taitoja. Tulevissa tutkimuksissa on syytä tutkia tarkemmin ja laajemmin yksilöllisten ja muiden tekijöiden merkitystä nuorten kriittisen nettilukemisen taidoissa ja niiden oppimisessa (ks. Anmarkrud ym., 2021).

Kontribuutiot ja implikaatiot

Väitöstutkimus kontribuoi laajan yhteiskunnallisen ongelman ratkaisemiseksi, sillä lisääntynyt epäluotettavan informaation leviäminen Internetissä vaikeuttaa tutkimukseen tai asiantuntijoiden välittämään informaatioon perustuvaa päätöksentekoa (cf. Ecker et al., 2022). Koska Internet on nykyään tärkeä tiedonlähde eri ikäisille ja leviävän misinformaation määrään on vaikea suoraan vaikuttaa, on tärkeä kehittää taitoja sekä tunnistaa epäluotettavampi informaatio että hyödyntää luotettavampaa informaatiota. Väitöstutkimuksen tulosten perusteella kriittisen nettilukemisen taitojen opettaminen on hyvä aloittaa jo peruskoulussa, joka tavoittaa kaikki erilaisista lähtökohdista tulevat lapset ja nuoret.

Tutkimuksella on merkittäviä metodologisia kontribuutioita ja implikaatioita, jotka edistävät aihepiirin tutkimusta. Ensinnäkin lukiolaisten nettitekstien luotettavuuden perusteluille kehitetyn pisteytystaulukon avulla voidaan huomioida sekä tärkeimpien arviointistrategioiden käyttö että perustelujen syvällisyys (vrt. Kiili ym., 2019). Lisäksi RCI-analyysillä (Jacobson & Truax, 1991) saadaan selville myös yksilötason muutoksia intervention aikana sekä se, millaisille oppilaille ja opiskelijoille interventio on hyödyllinen. Aiemmat interventiotutkimukset ovat pääsääntöisesti keskittyneet selvittämään intervention tehokkuutta vain ryhmätasolla (vrt. Brante & Strømsø, 2018), mutta RCI-analyysin käyttö auttaa kehittämään ja kohdentamaan kriittisen nettilukemisen opetusta erilaisille oppijoille sopivaksi. Väitöstutkimus myös osoitti, että yliopisto-opiskelijoilla kehitetty kysely Internet-spesifeistä episteemisistä uskomuksista (Bråten ym., 2019a) on toimiva myös lukioikäisillä ja kahdessa eri terveysaiheessa. Lisäksi tutkivan nettilukemisen tehtävän yhteydessä käytetyt täsmähakukoneet auttavat rajaamaan nettitekstivalintoja, mutta mahdollistavat samalla lähes autenttisen tiedonhaun ja nettitekstien valinnan.

Tutkimuksella on myös tärkeitä pedagogisia kontribuutioita ja implikaatioita. Väitöstutkimus osoitti, että kriittisen nettilukemisen taitojen opetus voidaan integroida koherentisti tutkivan nettilukemisen prosessin eri vaiheisiin (vrt. Leu ym., 2019). Interventioissa erilaiset opetusmenetelmät yhdistettiin kokonaisuudeksi niin, että oppitunnit sisälsivät oppilaille ja opiskelijoille ennustettavia rakenteita kuten esimerkiksi sen, että taitoja aina ensin mallinnettiin ja sen jälkeen harjoiteltiin. Erilaisista opetusmenetelmistä huolimatta eksplisiittinen opettaminen (vrt. Heijltjes ym., 2014; Marin & Helpern, 2011) näyttää olevan välttämätöntä kriittisen nettilukemisen taitojen oppimiselle. Lukiolaisten inter-

ventiossa onnistuneesti käytetty jaettu työskentelydokumentti edellyttää opiskelijoilta metakognitiivisten taitojen käyttöä (vrt. Stadtler & Bromme, 2007) eli yhteistä oppimisen ja työskentelyn suunnittelua sekä korostaa oppimisprosessia vain lopputuloksen sijaan. Digitaalinen työskentelydokumentti myös tekee ryhmäkohtaisen tutkivan nettilukemisen prosessin näkyväksi ja seurattavaksi opettajalle. Vaikka väitöstutkimuksen interventiot osoittautuivat toimiviksi, jatkossa on hyvä kiinnittää enemmän huomiota nuorten epistemisiin uskomuksiin, opetuksen eriyttämiseen, motivointiin sekä henkilökohtaisen palautteen antamiseen prosessin aikana.

REFERENCES

Abed, F., & Barzilai, S. (2023). Can students evaluate scientific YouTube videos? Examining students' strategies and criteria for evaluating videos versus webpages on climate change. *Journal of Computer Assisted Learning, 39*(2), 558–577. https://doi.org/10.1111/jcal.12762

Afflerbach, P. (Ed.). (2016). *Handbook of individual differences in reading: Reader, text, and context.* Routledge.

Anmarkrud, Ø., Bråten, I., Florit, E., & Mason, L. (2021). The role of individual differences in sourcing: a systematic review. *Educational Psychology Review, 34*(2), 749–792. https://doi.org/10.1007/s10648-021-09640-7

Applebee, A. N., Langer, J. A., Nystrand, M., & Gamoran, A. (2003). Discussion-based approaches to developing understanding: Classroom instruction and student performance in middle and high school English. *American Educational Research Journal, 40*(3), 685–730. https://doi.org/10.3102/00028312040003685

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

Aro, T., Viholainen, H., Koponen, T., Peura, P., Räikkönen, E., Salmi, P., Sorvo, R., & Aro, M. (2018). Can reading fluency and self-efficacy of reading fluency be enhanced with an intervention targeting the sources of self-efficacy? *Learning and Individual Differences, 67*, 53–66. https://doi.org/10.1016/j.lindif.2018.06.009

Barzilai, S., & Strømsø, H. I. (2018). Individual differences in multiple document comprehension. In Braasch, J. L., Bråten, I., & McCrudden, M. T. (Eds.). *Handbook of multiple source use* (pp. 99–116). Routledge.

Barzilai, S., Thomm, E., & Shlomi-Elooz, 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*, 737–766. https://doi.org/10.1007/s11251-015-9359-4

Binali, T., Tsai, C. C., & Chang, H. Y. (2021). University students' profiles of online learning and their relation to online metacognitive regulation and internet-specific epistemic justification. *Computers & Education, 175*, 104315. https://doi.org/10.1016/j.compedu.2021.104315

Braasch, J. L. G., Bråten, I., Strømsø, H. I., Anmarkrud, Ø., & Ferguson, L. E. (2013). Promoting secondary school students' evaluation of source features of multiple documents. *Contemporary Educational Psychology, 38*(3), 180–195. https://doi.org/10.1016/j.cedpsych.2013.03.003

Brand-Gruwel, S., & van Strien, J. L. H. (2018). Instruction to promote information-problem solving on the Internet in primary and secondary

education students: A systematic literature review. In Braasch, J. L., Bråten, I., & McCrudden, M. T. (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., Brandmo, C., Ferguson, L. E., & Strømsø, H. I. (2022). Epistemic justification in multiple document literacy: A refutation text intervention. *Contemporary Educational Psychology, 71,* 102122. https://doi.org/10.1016/j.cedpsych.2022.102122

Bråten, I., Brandmo, C., & Kammerer, Y. (2019a). A validation study of the Internet-Specific Epistemic Justification Inventory with Norwegian preservice teachers. *Journal of Educational Computing Research, 57*, 877–900. https://doi.org/10.1177/0735633118769438

Bråten, I., Brante, E. W., & Strømsø, 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., & Strømsø, H. I. (2019b). 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., Latini, N., & Haverkamp, Y. E. (2021). Predictors and outcomes of behavioral engagement in the context of text comprehension: When quantity means quality. *Reading and Writing, 35,* 687–711. https://doi.org/10.1007/s11145-021-10205-x

Bråten, I., McCrudden, M. T., Lund, E. S., Brante, E. W., & Strømsø, 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, & M. A. Britt (Eds.), *The Routledge handbook of discourse processes* (2nd ed., pp. 141–166). Routledge.

Britt, M. A., & Aglinskas, 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

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., Scharrer, L., Stadtler, M., Hömberg, J., & Torspecken, R. (2015). Is it believable when it's scientific? How scientific discourse style influences laypeople's resolution of conflicts. *Journal of Research in Science Teaching, 52*(1), 36–57. https://doi.org/10.1002/tea.21172

Bromme, R., & Goldman, S. R. (2014). The public's bounded understanding of science. *Educational Psychologist, 49*, 59–69. https://doi.org/10.1080/00461520.2014.921572

Brozo, W. G., Sulkunen, S., Shiel, G., Garbe, C., Pandian, A., & Valtin, R. (2014). Reading, gender, and engagement: Lessons from five PISA countries. *Journal Of Adolescent & Adult Literacy, 57*(7), 584–593. https://doi.org/10.1002/jaal.291

Cervetti, G., Pardales, M. J., & Damico, J. S. (2001). A tale of differences: Comparing the traditions, perspectives, and educational goals of critical reading and critical literacy. *Reading Online, 4*(9), 80–90. Available online: http://www.readingonline.org/articles/art_index.asp?HREF=/articles/cervetti/index.html

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.

Cheng, C. H., Bråten, I., Yang, F. Y., & Brandmo, C. (2021). Investigating structural relationships among upper-secondary school students' beliefs about knowledge, justification for knowing, and Internet-specific justification in the domain of science. *Journal of Research in Science Teaching, 58*(7), 980–1009. https://doi.org/10.1002/tea.21689

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

Choi, H. J., & Johnson, S. D. (2005). The effect of context-based video instruction on learning and motivation in online courses. *American Journal of Distance Education, 19*(4), 215–227. https://doi.org/10.1207/s15389286ajde1904_3

Christodoulou, S. A., & Diakidoy, I. A. N. (2020). The contribution of argument knowledge to the comprehension and critical evaluation of argumentative text. *Contemporary Educational Psychology, 63*, 101903. https://doi.org/10.1016/j.cedpsych.2020.101903

Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement, 20*(1), 37–46. https://doi.org/10.1177/001316446002000104

Cohen. J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Lawrence Erlbaum.

Coiro, J. (2011a). Predicting reading comprehension on the Internet: Contributions of offline reading skills, online reading skills, and prior knowledge. *Journal of Literacy Research, 43*(4), 352–392. https://doi.org/10.1177/1086296X11421979

Coiro, J. (2011b). Talking about reading as thinking: Modeling the hidden complexities of online reading comprehension. *Theory Into Practice, 50*(2), 107–115. https://doi.org/10.1080/00405841.2011.558435

Coiro, J., Coscarelli, C., Maykel, C., & Forzani, E. (2015). Investigating criteria that seventh graders use to evaluate the quality of online information. *Journal of Adolescent & Adult Literacy, 59,* 287–297. https://doi.org/10.1002/jaal.448

Coxe, S. (2018). Effect size measures for nonlinear count regression models. Poster presented at the *American Statistical Association Conference on Statistical Practice* 15.–17.2.2018 Portland: https://stefanycoxe.weebly.com/uploads/2/4/3/7/24372766/coxe_asa_csp_020618.pdf

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

De Hei, M., Strijbos, J.W., Sjoer, E., & Admiraal, W. (2016). Thematic review of approaches to design group learning activities in higher education: The development of a comprehensive framework. *Educational Research Review, 18,* 33–45. https://doi.org/10.1016/j.edurev.2016.01.001

de Jong, P. F., & van der Leij, A. (1999). Specific contributions of phonological abilities to early reading acquisition: Results from a Dutch latent variable longitudinal study. *Journal of Educational Psychology, 91*, 450–476. https://doi.org/10.1037/0022-0663.91.3.450

Ecker, U. K. H., Lewandowsky, S., Cook, J., Schmid, P., Fazio, L. K., Brashier, N., Kendeou, P., Vraga, E. K., & Amazeen, M. A. (2022). The psychological drivers of misinformation belief and its resistance to correction. *Nature Reviews Psychology 1,* 13–29. https://doi.org/10.1038/s44159-021-00006-y

Efron, B. (1987). Better bootstrap confidence intervals. *Journal of the American Statistical Association, 82*(397), 171–185.

The Council of the European Union. (2018). *EU Council recommendation on key competences for lifelong learning.* https://ec.europa.eu/education/education-in-the-eu/councilrecommendation-on-key-competences-for-lifelong-learning_en (Accessed 12 August 2019).

Finnish National Board on Research Integrity. (2019). *The ethical principles of research with human participants and ethical review in the human sciences in Finland.* TENK 3/2019. Available online: https://tenk.fi/sites/default/files/2021-01/Ethical_review_in_human_sciences_2020.pdf

Ferguson, L. E., Bråten, I., & Strømsø, H. I. (2012). Epistemic cognition when students read multiple documents containing conflicting scientific evidence: A think-aloud study. *Learning and Instruction, 22,* 103–120. https://doi.org/10.1016/j.learninstruc.2011.08.002

Freire, P. (1985). *The politics of education: Culture, power, and liberation.* Bergin & Garvey.

Frerejean, J., van Strien, J. L. H., Kirschner, P. A., & Brand-Gruwel, S. (2018). Effects of a modelling example for teaching information problem solving skills. *Journal of Computer Assisted Learning, 34*(6), 688–700. https://doi.org/10.1111/jcal.12276

Forzani, E. (2018). How well can students evaluate online science information? Contributions of prior knowledge, gender, socioeconomic status, and offline reading ability. *Reading Research Quarterly, 53*(4), 385–390. https://doi.org/10.1002/rrq.218

Forzani, E. (2020). A three-tiered framework for proactive critical evaluation during online inquiry. *Journal of Adolescent & Adult Literacy, 63*(4), 401–414. https://doi.org/10.1002/jaal.1004

Forzani, E., Corrigan, J., & Kiili, C. (2022). What does more and less effective internet evaluation entail? Investigating readers' credibility judgments across content, source, and context. *Computers in Human Behavior, 135*, 107359. https://doi.org/10.1016/j.chb.2022.107359

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

González-Ibañez, R., Gacitúa, D., Sormunen, E., & Kiili, C. (2017). Neurone: oNlinE inqUiRy experimentatiON systEm. *Proceedings of the Association for Information Science and Technology, 54*(1), 687–689. https://doi.org/10.1002/pra2.2017.14505401117

Greene, J. A., Azevedo, R., & Torney-Purta, J. (2008). Modeling epistemic and ontological cognition: Philosophical perspectives and methodological directions. *Educational Psychologist, 43,* 142–160. https://doi.org/10.1080/00461520802178458

Greene, J. A., Cartiff, B. M., Duke, R. F., & Deekens, V. M. (2019). A nation of curators. Educating students to be critical consumers and users of online information. In P. Kendeou, D. H. Robinson, & M. T. McCrudden (Eds.), *Misinformation and fake news in education* (pp. 187–206). Information Age Publishing.

Greenlaw, J. (2015). Deconstructing the metanarrative of the 21st Century Skills Movement. *Educational Philosophy and Theory, 47*(9), 894–903. http://dx.doi.org/10.1080/00131857.2015.1035156

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

Hahnel, C., Kroehne, U., Goldhammer, F., Schoor, C., Mahlow, N., & Artelt, C. (2019). Validating process variables of sourcing in an assessment of

multiple document comprehension. *British Journal of Educational Psychology, 89*(3), 524–537. https://doi.org/10.1111/bjep.12278
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*(1), 5–25. https://doi.org/10.1146/annurev-publhealth-040617-014128
Harju-Luukkainen, H., Vettenranta, J., Ouakrim-Soivio, N., & Bernelius, V. (2016). Differences between students' PISA reading literacy scores and grading for mother tongue and literature at school: A geostatistical analysis of the Finnish PISA 2009 data. *Education Inquiry, 7*(4), 463–479. https://doi.org/10.3402/edui.v7.29413
Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of Educational Research, 77*(1), 81–112. https://doi.org/10.3102/003465430298487
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
Hoeken, H. (2001). Anecdotal, statistical, and causal evidence: Their perceived and actual persuasiveness. *Argumentation, 15,* 425–437. https://doi.org/10.1023/A:1012075630523
Holopainen, L., Kairaluoma, L., Nevala, J., Ahonen, T., & Aro, M. (2004). *Lukivaikeuksien seulontatesti nuorille ja aikuisille [Dyslexia screening test for youth and adults].* Niilo Mäki Institute.
Iordanou, K., Muis, K. R., & Kendeou, P. (2019). Epistemic perspective and online epistemic processing of evidence: Developmental and domain differences. *The Journal of Experimental Education, 87*(4), 531–551. https://doi.org/10.1080/00220973.2018.1482857
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.
Kajamies, A. (2017). *Towards optimal scaffolding of low achievers' learning: Combining intertwined, dynamic, and multi-domain perspectives.* Doctoral dissertation, University of Turku, Finland: Painosalama Oy.
Kammerer, Y., Amann, D. G., & Gerjets, P. (2015). When adults without university education search the internet for health information: The roles of internet-specific epistemic beliefs and a source evaluation intervention. *Computers in Human Behavior, 48*, 297–309. https://doi.org/10.1016/j.chb.2015.01.045
Kammerer, Y., Bråten, I., Gerjets, P., & Strømsø, H. I. (2013). The role of internet-specific epistemic beliefs in laypersons' source evaluations and decisions during web search on a medical issue. *Computers in Human Behavior, 29*, 1193–1203. https://doi.org/10.1016/j.chb.2012.10.012
Kammerer, Y., Gottschling, S., & Bråten, I. (2021). The role of internet-specific justification beliefs in source evaluation and corroboration during web search on an unsettled socio-scientific issue. *Journal of Educational*

Computing Research, 59, 42–378. https://doi.org/10.1177/0735633120952731

Kammerer, Y., Kalbfell, E., & Gerjets, P. (2016a). Is this information source commercially biased? How contradictions between web pages stimulate the consideration of source information. *Discourse Processes, 53,* 430–456. https://doi.org/10.1080/0163853X.2016.1169968

Kammerer, Y., Meier, N., & Stahl, E. (2016b). 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

Kanniainen, L., Kiili, C., Tolvanen, A., Aro, M., & Leppänen, P. H. (2019). Literacy skills and online research and comprehension: struggling readers face difficulties online. *Reading and Writing, 32*(9), 2201–2222. https://doi.org/10.1007/s11145-019-09944-9

Kanniainen, L., Kiili, C., Tolvanen, A., Utriainen, J., Aro, M., Leu, D. J., & Leppänen, P. H. (2022). Online Research and Comprehension Performance Profiles Among Sixth-Grade Students, Including Those with Reading Difficulties and/or Attention and Executive Function Difficulties. *Reading Research Quarterly, 57*(4), 1213–1235. https://doi.org/10.1002/rrq.463

Kiili, C., W. Brante, E., Räikkönen, E., & Coiro, J. (2020). Citing as a sourcing practice: students' citing self-selected online sources in their essays (Las citas como práctica del uso de las fuentes: las citas de fuentes en línea seleccionados por los estudiantes en sus trabajos). *Journal for the Study of Education and Development, 43*(1), 174–209. https://doi.org/10.1080/02103702.2019.1690839

Kiili, C., Bråten, I., Strømsø, H. I., Hagerman, M. S., Räikkönen, E., & Jyrkiäinen, A. (2022a). Adolescents' credibility justifications when evaluating online texts. *Education and Information Technologies,* 1–30. https://doi.org/10.1007/s10639-022-10907-x

Kiili, C., Coiro, J., & Räikkonen, E. (2019). Students' evaluation of information during online inquiry: Working individually or in pairs. *Australian Journal of Language and Literacy*, *42*(3), 167–183. https://doi.org/10.1007/BF03652036

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. (2022b). 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., Laurinen, L., & Marttunen, M. (2008). Students evaluating Internet sources: From versatile evaluators to uncritical readers. *Journal of Educational Computing Research, 39*(1), 75–95. https://doi.org/10.2190%2FEC.39.1.e

Kiili, C., Leu, D. J., Marttunen, M., Hautala, J., & Leppänen, P. H. (2018a). Exploring early adolescents' evaluation of academic and commercial online resources related to health. *Reading and Writing, 31*(3), 533–557. https://doi.org/10.1007/s11145-017-9797-2

Kiili, C., Leu, D. J., Utriainen, J., Coiro, J., Kanniainen, L., Tolvanen, A., Lohvansuu, K., & Leppänen, P. H. (2018b). Reading to learn from online information: Modeling the factor structure. *Journal of Literacy Research, 50*(3), 304–334. https://doi.org/10.1177%2F1086296X18784640

Kiili, C., Räikkönen, E., Bråten, I., Strømsø, H. I., & Hagerman, M. S. (2023). Examining the structure of credibility evaluation when sixth graders read online texts. *Journal of Computer Assisted Learning* (Published online), https://doi.org/10.1111/jcal.12779

Kingsley, T. L., Cassady, J. C., & Tancock, S. M. (2015). Successfully promoting 21st century online research skills: Interventions in 5th-grade classrooms. *Reading Horizons, 54*(2), 91–135. https://scholarworks.wmich.edu/reading_horizons/vol54/iss2/5

Kintsch, W. (1988). The use of knowledge in discourse processing: A construction-integration model. *Psychological Review, 95*, 163–182.

Kohnen, A. M., & Mertens, G. E. (2019). "I'm always kind of double-checking": Exploring the information-seeking identities of expert generalists. *Reading Research Quarterly, 54,* 279–297. https://doi.org/10.1002/rrq.245

Kuhlthau, C. C. (1991). Inside the search process: Information seeking from the user's perspective. *Journal of the American Society for Information Science, 42*(5), 361–371. https://doi.org/10.1002/(SICI)1097-4571(199106)42:5%3C361::AID-ASI6%3E3.0.CO;2-#

Kulju, P., Hämäläinen, E. K., Mäkinen, M., Räikkönen, E., & Kiili, C. (under review). Pre-service Teachers Evaluating Online Texts About Learning Styles: There Is Room for Improvement in Justifying the Credibility. Pre-print version available.

Kullberg, N., Kiili, C., Bråten, I., González-Ibáñez, R., & Leppänen, P. H. T. (2023). Sixth graders' selection and integration when writing from multiple online texts. *Instructional Science 51*, 39–64. https://doi.org/10.1007/s11251-022-09613-5

Larson, A. A., Britt, M. A., & Kurby, C. A. (2009). Improving students' evaluation of informal arguments. *The Journal of Experimental Education, 77*(4), 339–366. https://doi.org/10.3200/JEXE.77.4.339-366

Leino, K., Rautopuro, J., & Kulju, P. (2021). *Lukutaito – tie tulevaisuuteen: PISA 2018 Suomen pääraportti. [Literacy - a guide to the future: PISA 2018 main report of Finland].* Kasvatusalan tutkimuksia, (82).

Leu, D. J., Forzani, E., Rhoads, C., Maykel, C., Kennedy, C., & Timbrell, N. (2015). The new literacies of online research and comprehension: Rethinking the reading achievement gap. *Reading Research Quarterly, 50*(1), 37–59. https://doi.org/10.1002/rrq.85

Leu, D. J., Kinzer, C. K., Coiro, J., Castek, J., & Henry, L. A. (2019). New literacies: A dual level theory of the changing nature of literacy, instruction, and assessment. In D. E. Alvermann, N. J. Unrau, M. Sailors, &

R. B. Ruddell (Eds.), *Theoretical models and processes of literacy* (7th ed., pp. 319–346). Taylor & Francis.
Lewandowsky, S., Ecker, U. K., & Cook, J. (2017). Beyond misinformation: Understanding and coping with the "post-truth" era. *Journal of Applied Research in Memory and Cognition*, *6*(4), 353–369. https://doi.org/10.1016/j.jarmac.2017.07.008
Lewandowsky, S., Stritzke, W. G. K., Freund, A. M., Oberauer, K. & Krueger, J. I. (2013). Misinformation, disinformation, and violent conflict: from Iraq and the war on terror to future threats to peace. *American Psychologist, 68*, 487–501. https://psycnet.apa.org/doi/10.1037/a0034515
List, A., & Alexander, P. A. (2017). Cognitive affective engagement model of multiple source use. *Educational Psychologist, 52,* 182–199. https://doi.org/10.1080/00461520.2017.1329014
List, A., Alexander, P. A., & Stephens, L. A. (2017). Trust but verify: Examining the association between students' sourcing behaviors and ratings of text trustworthiness. *Discourse Processes, 54*(2), 83–104. https://doi.org/10.1080/0163853X.2016.1174654
Liu, Z., & Huang, X. (2008). Gender differences in the online reading environment. *Journal of Documentation, 64*(4), 616–626. https://doi.org/10.1108/00220410810884101
Lucassen, T., Muilwijk, R., Noordzij, M.L., & Schraagen, J.M. (2013). Topic familiarity and information skills in online credibility evaluation. *Journal of the American Society for Information Science and Technology, 64*(2), 254–264. https://doi.org/10.1002/asi.22743
Lucassen, T., & Schraagen, J.M. (2011). Factual accuracy and trust in information: The role of expertise. *Journal of the American Society for Information Science and Technology, 62*(7), 1232–1242. https://doi.org/10.1002/asi.21545
Macedo-Rouet, M., Braasch, J. L. G., Britt, M. A., & Rouet, J.-F. (2013). Teaching fourth and fifth graders to evaluate information sources during text comprehension. *Cognition and Instruction, 31*(2), 204–226. https://doi.org/10.1080/07370008.2013.769995
Macedo-Rouet, M., Potocki, A., Scharrer, L., Ros, C., Stadtler, M., Salmeròn, L., & Rouet, J.-F. (2019). How good is this page? Benefits and limits of prompting on adolescents' evaluation of web information quality. *Reading Research Quarterly, 54*(3), 299–321. https://doi.org/10.1002/rrq.241
Macedo-Rouet, M., Salmerón, L., Ros, C., Pérez, A., Stadtler, M., & Rouet, J.-F. (2020). Are frequent users of social network sites good information evaluators? An investigation of adolescents' sourcing abilities. *Journal for the Study of Education and Development, 43*(1), 101–138. https://doi.org/10.1080/02103702.2019.1690849
Mahlow, N., Hahnel, C., Kroehne, U., Artelt, C., Goldhammer, F., & Schoor, C. (2020). More than (single) text comprehension? – On university students' understanding of multiple documents. *Frontiers in Psychology, 11*, 562450. https://doi.org/10.3389/fpsyg.2020.562450

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

Marôco, J. (2021). What makes a good reader? Worldwide insights from PIRLS 2016. *Reading and Writing, 34*(1), 231–272. https://doi.org/10.1007/s11145-020-10068-8

Marttunen, M., Salminen, T., & Utriainen, J. (2021). Student evaluations of the credibility and argumentation of online sources. *The Journal of Educational Research, 114*(3), 294–305. https://doi.org/10.1080/00220671.2021.1929052

McCarthy, K. S., & McNamara, D. S. (2021). The multidimensional knowledge in text comprehension framework. *Educational Psychologist, 56*(3), 196–214. https://doi.org/10.1080/00461520.2021.1872379

McCrudden, M. T. (2018). Text relevance and multiple-source use. In J. L. G. Braasch, I. Bråten, & M. T. McCrudden (Eds.), *Handbook of multiple source use* (pp. 168–183). Routledge.

McGrew, S., Breakstone, J., Ortega, T., Smith, M., & Wineburg, S. (2018). Can students evaluate online sources? Learning from assessments of civic online reasoning. *Theory & Research in Social Education, 46*(2), 165–193. https://doi.org/10.1080/00933104.2017.1416320

McGrew, S. (2020). Learning to evaluate: An intervention in civic online reasoning. *Computers & Education, 145,* 103711. https://doi.org/10.1016/j.compedu.2019.103711

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

McGrew, S., & Byrne, V. L. (2022). Conversations after lateral reading: Supporting teachers to focus on process, not content. *Computers & Education, 85*, 104519. https://doi.org/10.1016/j.compedu.2022.104519

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

Means, M. L., & Voss, J. F. (1996). Who reasons well? Two studies of informal reasoning among children of different grade, ability, and knowledge levels. *Cognition and Instruction, 14*(2), 139–178. https://doi.org/10.1207/s1532690xci1402_1

Murphy, P. K., & Alexander, P. A. (2004). Persuasion as a dynamic, multidimensional process: An investigation of individual and intraindividual differences. *American Educational Research Journal, 41*, 337–363.

Muthén, L., & Muthén, B. O. (1998–2017). *Mplus user's guide* (8th ed). Muthén & Muthén.

National Core Curriculum for Basic Education. (2014). Finnish National Agency for Education (OPH). (Publication No. 2016:5).

National Core Curriculum for General Upper Secondary Education. (2019). Finnish National Agency for Education (OPH). (Publication No. 2019:2c).

New London Group. (1996). The pedagogy of multiliteracies: Designing social futures. *Harvard Educational Review, 66*(1), 60–93. https://doi.org/10.17763/haer.66.1.17370n67v22j160u

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

Paakkari, L., Okan, O., & Torppa, M. (2022). "Missed" Information: A Moral Failing that Erodes Efforts to Tackle the COVID-19 Pandemic. *International Journal of Public Health*, 126. https://doi.org/10.3389/ijph.2021.1604667

Palsa, L., & Ruokamo, H. (2015). Behind the concepts of multiliteracies and media literacy in the renewed Finnish core curriculum: A systematic literature review of peer-reviewed research. *Seminar. net (11)*2. https://doi.org/10.7577/seminar.2354

Pan, B., Hembrooke, H., Joachims, T., Lorigo, L., Gay, G., & Granka, L. (2007). In Google we trust: Users' decisions on rank, position, and relevance. *Journal of Computer-Mediated Communication, 12*, 801–823. https://doi.org/10.1111/j.1083-6101.2007.00351.x

Paul, J., Macedo-Rouet, M., Rouet, J. F., & Stadtler, M. (2017). Why attend to source information when reading online? The perspective of ninth grade students from two different countries. *Computers & Education, 113,* 339–354. https://doi.org/10.1016/j.compedu.2017.05.020

Pavlik, J. (2022). The Russian War in Ukraine and the Implications for the News Media. *Athens Journal of Mass Media and Communications, 8*, 1–17. https://doi.org/10.30958/ajmmc.X-Y-Z

Pedaste, M., Mäeots, M., Siiman, L. A., De Jong, T., Van Riesen, S. A., Kamp, E. T., Manoli, C. C., Zacharia, Z. C., & Tsourlidaki, E. (2015). Phases of inquiry-based learning: Definitions and the inquiry cycle. *Educational Research Review*, *14*, 47–61. https://doi.org/10.1016/j.edurev.2015.02.003

Pérez, A., Potocki, A., Stadtler, M., Macedo-Rouet, M., Paul, J., Salmerón, L., & Rouet, J. F. (2018). Fostering teenagers' assessment of information reliability: Effects of a classroom intervention focused on critical source dimensions. *Learning and Instruction, 58*, 53–64. 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., Stadtler, M., Salmeron, L., & Rouet, J.-F. (2020). The development of source evaluation skills during adolescence: Exploring different levels of source processing and their relationships. *Journal for the Study of Education and Development, 43*, 19–59. https://doi.org/10.1080/02103702.2019.1690848

Potter, W. J. (2022). Analysis of definitions of media literacy. *Journal of Media Literacy Education, 14*(2), 27–43. https://doi.org/10.23860/JMLE-2022-14-2-3

Rasi, P., Kangas, M., & Ruokamo, H. (2019). Promoting multiliteracy in the Finnish educational system. *Introduction to the Finnish educational system*, 97–111. https://doi.org/10.1163/9789004394278_007

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

Roozenbeek, J., Schneider, C. R., Dryhurst, S., Kerr, J., Freeman, A. L. J., Recchia, G., van der Bles, A. M., & van der Linden, S. (2020). Susceptibility to misinformation about COVID-19 around the world. *Royal Society Open Science. 7*, 201199. https://doi.org/10.1098/rsos.201199

Rouet, J. F. (2006). *The skills of document use: From text comprehension to Web-based learning.* Routledge.

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

Salmerón, L., Gil, L., & Bråten, I. (2018a). 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

Salmerón, L., Sampietro, A., & Delgado, P. (2020). Using Internet videos to learn about controversies: Evaluation and integration of multiple and multimodal documents by primary school students. *Computers & Education, 148*, 103796. https://doi.org/10.1016/j.compedu.2019.103796

Salmerón, L., Strømsø, H. I., Kammerer, Y., Stadtler, M., & van den Broek, P. (2018b). Comprehension processes in digital reading. In M. Barzillai, J. Thomson, S. Schroeder, & P. van den Broek (Eds.), *Learning to read in a digital world* (pp. 91–120). Studies in written language and literacy 17. John Benjamins Publishing Company.

Sandoval, W. A., Sodian, B., Koerber, S., & Wong, J. (2014). Developing children's early competencies to engage with science. *Educational Psychologist, 49*, 39–152. https://doi.org/10.1080/00461520.2014.917589

Saux, G., Britt, M. A., Vibert, N., & Rouet, J. F. (2021). Building mental models from multiple texts: How readers construct coherence from inconsistent sources. *Language and Linguistics Compass, 15*(3), 12409. https://doi.org/10.1111/lnc3.12409

Scharrer, L., Stadtler, M., & Bromme, R. (2019). Judging scientific information: Does source evaluation prevent the seductive effect of text easiness? *Learning and Instruction, 63*, 101215. https://doi.org/10.1016/j.learninstruc.2019.101215

Shute, V. J. (2008). Focus on formative feedback. *Review of Educational Research, 78*(1), 153–189. https://doi.org/10.3102/0034654307313795

Sinatra, G. M., & Lombardi, D. (2020). Evaluating sources of scientific evidence and claims in the post-truth era may require reappraising plausibility judgments. *Educational Psychologist, 55,* 120–131. https://doi.org/10.1080/00461520.2020.1730181

Sormunen, E., Erdmann, N., Otieno, S. C., Mikkilä-Erdmann, M., Laakkonen, E., Mikkonen, T., Hossain, M. A., González-Ibañez, R., Quintanilla-Gatica, M., Leppänen, P. H. T., & Vauras, M. (2021). How do gender, Internet activity and learning beliefs predict sixth-grade students' self-efficacy beliefs in and attitudes towards online inquiry? *Journal of Information Science*. Available online: https://doi.org/10.1177/01655515211043708

Sormunen, E., González-Ibañez, R., Kiili, C., Leppänen, P. H. T., Mikkilä-Erdmann, M., Erdmann, N., et al. (2018). A performance-based test for assessing students' online inquiry competences in schools. In S. Kurbanoglu, J. Boustany, S. Spiranec, E. Grassian, D. Mizrachi, & L. Roy (Eds.), *Information literacy in the workplace. ECIL, 2017. Communications in computer and information science, 810* (pp. 673–682).

Sparks, J. R., van Rijn, P. W., & Deane, P. (2021). Assessing source evaluation skills of middle school students using learning progressions. *Educational Assessment, 26*(4), 213–240. https://doi.org/10.1080/10627197.2021.1966299

Stadtler, M., & Bromme, R. (2007). Dealing with multiple documents on the WWW: The role of metacognition in the formation of documents models. *International Journal of Computer-Supported Collaborative Learning, 2,* 191–210. https://doi.org/10.1007/s11412-007-9015-3

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.

Stewart, E. (2021). Detecting fake news: Two problems for content moderation. *Philosophy & Technology, 34*(4), 923–940. https://doi.org/10.1007/s13347-021-00442-x

Strømsø, H. I., Bråten, I., & Britt, M. A. (2011). Do students' beliefs about knowledge and knowing predict their judgement of texts' trustworthiness? *Educational Psychology, 31*(2), 177–206. https://doi.org/10.1080/01443410.2010.538039

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/opiskt/2020/opiskt_2020_2021-06-10_tau_001_fi.html
Suomen virallinen tilasto [Official Statistics of Finland], (2022). *Esi- ja peruskouluopetus. [Pre-primary and comprehensive school education.]* Tilastokeskus. http://www.stat.fi/til/pop/
Tsai, M. J., Wu, A. H., Bråten, I., & Wang, C. Y. (2022). What do critical reading strategies look like? Eye-tracking and lag sequential analysis reveal attention to data and reasoning when reading conflicting information. *Computers & Education, 187*, 104544. https://doi.org/10.1016/j.compedu.2022.104544
Tsamakis, K., Tsiptsios, D., Stubbs, B., Ma, R., Romano, E., Mueller, C., Ahmad, A., Triantafyllis, A. S., Tsitas, G., & Dragioti, E. (2022). Summarising data and factors associated with COVID-19 related conspiracy theories in the first year of the pandemic: a systematic review and narrative synthesis. *BMC psychology*, *10*(1), 244. https://doi.org/10.1186/s40359-022-00959-6
Ulyshen, T. Z., Koehler, M. J., & Gao, F. (2015). Understanding the connection between epistemic beliefs and Internet searching. *Journal of Educational Computing Research, 53*(3), 345–383. https://doi.org/10.1177/0735633115599604
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
von der Mühlen, S., Richter, T., Schmid, S., Schmidt, E. M., & Berthold, K. (2016). Judging the plausibility of arguments in scientific texts: A student–scientist comparison. *Thinking & Reasoning, 22*(2), 221–249. https://doi.org/10.1080/13546783.2015.1127289
Walraven, A., Brand-Gruwel, S., & Boshuizen, H. P. (2009). How students evaluate information and sources when searching the World Wide Web for information. *Computers & Education, 52*(1), 234–246. https://doi.org/10.1016/j.compedu.2008.08.003
Walraven, A., Brand-Gruwel, S., & Boshuizen, H. P. A. (2013). Fostering students' evaluation behaviour while searching the Internet. *Instructional Science, 41*(1), 125–146. https://doi.org/10.1007/s11251-012-9221-x
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
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
White, M. D., & Marsh, E. E. (2006). Content analysis: A flexible methodology. *Library Trends, 55*(1), 22–45. https://doi.org/10.1353/lib.2006.0053
Wuyckens, G., Landry, N., & Fastrez, P. (2022). Untangling media literacy, information literacy, and digital literacy: a systematic meta-review of core

concepts in media education. *Journal of Media Literacy Education, 14*(1), 168–182. https://doi.org/10.23860/JMLE-2022-14-1-12

Yew, E. H., & Goh, K. (2016). Problem-based learning: An overview of its process and impact on learning. *Health Professions Education*, *2*(2), 75–79. https://doi.org/10.1016/j.hpe.2016.01.004

Zhang, S., & Duke, N. K. (2011). The impact of instruction in the WWWDOT framework on student's disposition and ability to evaluate web sites as sources of information. *The Elementary School Journal, 112*(1), 132–154. https://doi.org/10.1086/660687

Zurkowski, P. G. (1974). The information service environment: Relationships and priorities (Related Paper No. 5).

ORIGINAL PAPERS

I

PROMOTING SIXTH GRADERS' CREDIBILITY EVALUATION OF WEB PAGES: AN INTERVENTION STUDY

by

Elina K. Hämäläinen, Carita Kiili, Miika Marttunen, Eija Räikkönen,
Roberto González-Ibáñez, & Paavo H. Leppänen, 2020

Computers in Human Behavior, 110, 106372.

https://doi.org/10.1016/j.chb.2020.106372

Computers in Human Behavior 110 (2020) 106372

Contents lists available at ScienceDirect

Computers in Human Behavior

journal homepage: http://www.elsevier.com/locate/comphumbeh

Full length article

Promoting sixth graders' credibility evaluation of Web pages: An intervention study

Elina K. Hämäläinen [a,*], Carita Kiili, PhD [b], Miika Marttunen, Professor [c], Eija Räikkönen, PhD [a], Roberto González-Ibáñez, Associate Professor [d], Paavo H.T. Leppänen, Professor [e]

[a] Faculty of Education and Psychology, University of Jyväskylä, P.O. Box 35, 40014, University of Jyväskylä, Finland
[b] Department of Education, University of Oslo, Norway and Faculty of Education and Culture, Tampere University, P.O. Box 700, 33014, Tampere University, Finland
[c] Department of Education, University of Jyväskylä P.O. Box 35, 40014, University of Jyväskylä, Finland
[d] Departamento de Ingeniería Informática, Universidad de Santiago de Chile, Chile
[e] Department of Psychology, University of Jyväskylä P.O. Box 35, 40014, University of Jyväskylä, Finland

ARTICLE INFO

Keywords:
Evaluation
Internet
Intervention
Adolescents
Online reading

ABSTRACT

This study investigated whether a teacher-led intervention program on online inquiry improved sixth graders' performance in a credibility evaluation task. Students ($N = 342$) were divided into two conditions, an intervention group (190 students) and a control group (152 students). The intervention program (21 × 45 min lessons) was implemented during a six-week course as a part of normal schoolwork. The program included explicit teaching of online inquiry skills: searching for information (3 lessons), evaluating credibility of information (3 lessons), and synthesizing information (3 lessons). In addition, the skills taught were applied in two online inquiry projects comprising 12 lessons in total. The control group received business-as-usual teaching. Students' performance in the credibility evaluation task was measured before and after the program by pre- and post-tests. In these tests, students evaluated three Web pages dealing with two topics (Computer Gaming or Reading on Screen) varying in their perspectives and argumentation. Students rated the credibility of each Web page and justified their ratings. Topic order was counterbalanced in both conditions. The background variables (Pre-test scores, Reading comprehension, Reading fluency, Gender, Topic order, and Test order in the pre-test) were controlled for in the multilevel negative binomial regression analysis. The results showed that the intervention program helped students better justify their credibility ratings by reference to source features but not to the argumentation or other aspects of the content compared to controls. Instructional implications of the findings are discussed.

1. Introduction

Given that the Internet, and the variable quality of the resources it offers, has become a dominant information channel, educating consumers to be critical of online information has also become an important goal of basic education (EU, 2018; Finnish National Board of Education, 2016). For even the youngest students, the Internet is a crucial information resource (Livingstone, Mascheroni, & Staksrud, 2018). Therefore, instruction targeted at fostering critical evaluation skills should take place already in the early school years (Leu et al., 2015). Recent research findings that adolescents read and use online information in a rather superficial and uncritical manner indicates that supporting students to become critical consumers of online information is of paramount importance (Coiro, Coscarelli, Maykel, & Forzani, 2015; Macedo-Rouet et al., 2019).

Unfortunately, school traditions and teaching practices often continue to emphasize the reading of single texts, and hence one-sided learning content, instead of the processing of multiple documents and materials presenting conflicting views (Macedo-Rouet, Braasch, Britt, & Rouet, 2013; Paul, Macedo-Rouet, Rouet, & Stadtler, 2017). The latter approach would better prepare students for the demands of working with unfiltered online information. Along with multiple document literacy practices, students need explicit instruction in how to evaluate the credibility of information when working with multiple information

* Corresponding author.
E-mail addresses: elina.k.hamalainen@jyu.fi (E.K. Hämäläinen), carita.kiili@tuni.fi (C. Kiili), miika.marttunen@jyu.fi (M. Marttunen), eija.m.raikkonen@jyu.fi (E. Räikkönen), roberto.gonzalez.i@usach.cl (R. González-Ibáñez), paavo.ht.leppanen@jyu.fi (P.H.T. Leppänen).

https://doi.org/10.1016/j.chb.2020.106372
Received 14 August 2019; Received in revised form 31 March 2020; Accepted 2 April 2020
Available online 9 April 2020

resources on the Web. While increasing efforts have been made to develop intervention programs to promote students' evaluation skills (e.g., Bråten, Brante, & Strømsø, 2019; Pérez et al., 2018), teacher-led interventions remain very rare among elementary school students (see reviews by Brand-Gruwel & van Strien, 2018; Brante & Strømsø, 2018).

To address these requirements, we developed and tested a teacher-led intervention program for elementary school students (sixth graders) in which evaluation of the credibility of Web pages was contextualized as a part of an online inquiry and multiple document reading. The intervention program focused on the central credibility evaluation skills that only a few adolescents are able to apply when reading and learning online (e.g., Forzani, 2018; Kiili, Leu, Marttunen, Hautala, & Leppänen, 2018).

1.1. Credibility evaluation

Credibility evaluation is crucial when reading single texts (Pressley & Afflerbach, 1995) but even more critical when reading multiple texts on paper (Perfetti, Rouet, & Britt, 1999) or online (Brand-Gruwel, Wopereis, & Walraven, 2009; Leu et al., 2015). Evaluating the quality of information presented in multiple texts refers to evaluating the credibility of the source (source-based evaluation) and the credibility of the ideas presented in the text (content-based evaluation) both within and across texts (Forzani, 2019; Stadtler & Bromme, 2014).

The documents model framework (Britt, Richter, & Rouet, 2014; Perfetti et al., 1999) highlights the importance of information about the source of texts, also often termed *source features,* when readers evaluate and process multiple texts. Source features can include the authors' expertise, credentials, affiliation and motives, and document type and date. When readers engage in *sourcing* they identify and draw on source features in different phases of reading to predict, interpret, and evaluate and use document content in relation to the reading task (Brante & Strømsø, 2018; Bråten, Stadtler, & Salmerón, 2018b). However, recent studies have shown that integration of source information in written products is challenging for secondary school students (Kiili, Brante, Räikkönen, & Coiro, 2020b; Pérez et al., 2018) and especially for students at the elementary school level (Kiili, Bråten, Kullberg, & Leppänen, 2020a).

Content-based evaluation can refer to evaluation of the argumentative purpose of the text (Mateos et al., 2018) and the accuracy of the ideas presented in the text, including evaluation of the argumentation used, i.e., the author's claims, reasons and evidence, and explanations given (Forzani, 2019). In evaluating argumentation, readers judge whether arguments are supported with valid reasons (Larson, Britt, & Kurby, 2009), whether both sides of an issue are considered (Means & Voss, 1996) and the quality of the evidence (e.g., single experience, research) presented (Hoeken, 2001). To evaluate the content quality, readers can also corroborate information using their prior knowledge or other texts (Forzani, 2019).

It should be noted that source features and content should not be perceived as separate entities when evaluating the credibility of Web pages (Paul et al., 2017). Two models building on Documents Model, the Discrepancy-Induced Source Comprehension (D-ISC) Model (Bråten & Braasch, 2018) and the Content-Source Integration (CSI) Model (Stadtler & Bromme, 2014) explain the reciprocal relationship between evaluation of knowledge claims and evaluation of sources when readers encounter conflicting scientific or socio-scientific information. According to D-ISC Model, readers' attention to sources increases when different sources provide conflicting information about the issue under examination. In these types of situations, linking sources to conflicting pieces of information helps readers to organize conflicting views that further enable them to build a coherent mental representation on the issue. Furthermore, paying attention to the source features can help readers to understand why different Web pages represent diverse perspectives or positions (Brante & Strømsø, 2018).

Further, CSI Model (Stadtler & Bromme, 2014) explains how readers may resolve conflicts when evaluating competing knowledge claims. Readers can evaluate the validity of claims by relying on their own understanding of the issue or evaluating the source of information. Particularly in situations where readers do not have sufficient prior knowledge, they may prefer to rely on their evaluations of author's expertise and intentions (see also Bråten, McCrudden, Lund, Brante, & Strømsø, 2018a). Stadtler and Bromme (2014) stress that often these two ways of evaluation tend to complement each other.

1.2. Previous intervention research

Given the importance of the ability to evaluate the credibility of Web pages and the limited nature of students' skills for doing so (Kiili et al., 2018; Macedo-Rouet et al., 2019), several studies have tested instructional methods designed to improve these skills. Recent reviews on sourcing (Brante & Strømsø, 2018) and online inquiry interventions (Brand-Gruwel & van Strien, 2018) show that most studies have been conducted among students in upper secondary or higher education institutions (e.g., Bråten et al., 2019). Notably, fewer interventions with a control group have been conducted among younger students. Most of the interventions carried out at the lower secondary level (mean student age 15 years) have reported positive results (e.g., Argelagós & Pifarré, 2012; Mason, Junyent, & Tornatora, 2014; Pérez et al., 2018) whereas those at the elementary school level (mean student age 10 years) have shown more varying success in enhancing students' credibility evaluation skills (e.g., Kingsley, Cassady, & Tancock, 2015; Macedo-Rouet et al., 2013; Zhang & Duke, 2011). These interventions with younger students have differed in several important respects.

First, the measures used to evaluate the efficacy of the instructional methods employed have differed. While some studies have applied credibility scales to single Web pages or have rank-ordered or rated the credibility of Web pages (Kammerer, Meier, & Stahl, 2016; Mason et al., 2014), others have asked students not only to rate but also to justify their credibility ratings or Web page evaluations in their own words (e.g., Kroustallaki, Kokkinaki, Sideridis, & Simos, 2015; Walraven, Brand-Gruwel, & Boshuizen, 2013; Zhang & Duke, 2011). While the formulation of justifications can be a demanding task for young students, it may also reveal the criteria they are able to apply when evaluating Web pages. It is of crucial importance that the evaluation tasks are selected so that they are neither too demanding for all the students (Kingsley et al., 2015; Zhang & Duke, 2011) nor too easy for the more skilled students (Macedo-Rouet et al., 2013).

Second, interventions have varied in length, content, and whether they have been led by a teacher or/and a researcher. Teacher-led interventions have lasted for several lessons, often including all phases of online inquiry: searching, evaluating, synthesizing, and communicating information to others (Argelagós & Pifarré, 2012; Kingsley et al., 2015; Walraven et al., 2013). In contrast, researcher-led interventions have typically lasted for one lesson only and focused only on specific aspects of online inquiry such as evaluating the credibility of Web pages (Kammerer et al., 2016; Macedo-Rouet et al., 2013; Mason et al., 2014). Owing, for example, to teachers' problems in implementing intervention programs (Walraven et al., 2013), teacher-led, long-lasting interventions have not invariably shown improvement in students' credibility evaluation skills (Brante & Strømsø, 2018). In turn, the effects of researcher-led interventions on students' credibility evaluation skills have been either positive (e.g., Mason et al., 2014) or partially positive (Kammerer et al., 2016; Macedo-Rouet et al., 2013).

Even though brief researcher-led interventions have shown positive outcomes in students' credibility evaluation skills (Bråten et al., 2018a; Mason et al., 2014), it is nevertheless important to involve teachers in the implementation of intervention programs to learn how instructional methods work in regular classrooms. In the present study, the teaching of credibility evaluation skills was embedded in the teaching program as one component of an online inquiry intervention designed by the researchers but implemented by regular classroom teachers.

1.3. Instruction for credibility evaluation

In this section, we introduce several instructional features that we applied in designing the present intervention program: modeling (Coiro, 2011; Davey, 1983), cognitive and metacognitive prompts (Berthold, Nückles, & Renkl, 2007; Quintana, Zhang, & Krajcik, 2005), discussions (Applebee, Langer, Nystrand, & Gamoran, 2003; Teasley, 1995), and eliciting sourcing behaviour by presenting sources providing conflicting information (Brante & Strømsø, 2018; Pérez et al., 2018).

As recent research has shown, many young students have limited credibility evaluation skills (Coiro et al., 2015; Kiili et al., 2018). One useful way to introduce students to different evaluation strategies is to *model* them. For example, a teacher may model the processes of skilled evaluators interacting with Web pages (Coiro, 2011; Davey, 1983) or display a video showing a person applying evaluation strategies (van Gog, Verveer, & Verveer, 2014; van Wermeskerken & van Gog, 2017). Modeling seems to be the most effective method for novice learners, and it seems to be important that it is followed by active engagement with the modeled strategies (Frerejean, van Strien, Kirschner, & Brand-Gruwel, 2018). Students can analyze the target skills by themselves (Fisher & Frey, 2015), contrast their own approach with that of an expert (Frerejean et al., 2018), or compare the performance of two models in which the evaluation skills are at different levels (Bråten et al., 2019). In addition, to strengthen and automate the modeled and analyzed skills, they must be practiced (Fisher & Frey, 2015; Frerejean et al., 2018).

Research has also shown that in many cases students do not spontaneously evaluate Web pages (Paul et al., 2017). These students would probably benefit from *prompting*, a method that seeks to direct their attention to important aspects of learning and to stimulate cognitive activities that students do not otherwise execute spontaneously (Quintana et al., 2005; Zhang, Hsu, Wang, & Ho, 2015). For example, cognitive prompts can be used to help students pay attention to the source features (e.g., Bråten et al., 2018a; Paul et al., 2017) or remind them to stop and consider the quality of the content (e.g., Britt & Aglinskas, 2002) of the Web pages they are reading. These prompts can be embedded in digital tools (Kiili, Coiro, & Räikkönen, 2019; Quintana et al., 2005) or worksheets (Kammerer et al., 2016; Zhang & Duke, 2011).

According to Pérez et al. (2018), written prompts alone are not enough to engage students in productive sourcing behaviour, and hence additional group *discussions* are needed. Furthermore, discussions are crucial in facilitating students' awareness of the connections between different source features and the reliability of Web pages (Bråten et al., 2019; Macedo-Rouet et al., 2013; Pérez et al., 2018). It also seems important that such discussions in the classrooms are structured, focus on the most important aspects to be learned, and carefully implemented with sufficient time resources (Walraven, Brand-Gruwel, & Boshuizen, 2010; 2013).

As well as models and prompts, it is important that teachers provide students with Web pages that elicit use of the modeled and prompted evaluation strategies. For example, providing *multiple partly conflicting Web pages* can trigger sourcing activities (Bråten et al., 2019; Paul et al., 2017) and improve learning outcomes related to the evaluation of Web pages (Mason et al., 2014; Pérez et al., 2018). The evaluation of multiple conflicting Web pages encourages students to pay more attention to source features compared to the evaluation of a single page (cf. Macedo-Rouet et al., 2013) or multiple pages presenting the same point of view (Brante & Strømsø, 2018). Further, the use of pre-selected Web pages, instead of using student-selected pages from open-Web sources, enables deeper-level discussions in which students can compare their evaluations of the same pages with their peers (Brante & Strømsø, 2018).

Applying the instructional features outlined above, this study examined whether a teacher-led intervention improved sixth graders' performance on an online credibility evaluation task.

1.4. The present study

The present study extends the previous intervention research on students' credibility evaluation skills by examining whether sixth graders improved in their online credibility evaluation by a teacher-led intervention. The performance of the intervention group was compared with that of the control group receiving business-as-usual -teaching. Furthermore, we also examined how students' credibility evaluations were reflected in their written products.

The specific research questions were:

1. How did the sixth graders evaluate the credibility of Web pages?
2. Did the teacher-led intervention lead to improvement in the sixth graders' performance on an online credibility evaluation task compared to the control group?
3. How were students' credibility evaluations reflected in their written products? Did the teacher-led intervention result in increase of students' use of justifications for credibility in their written products?

In terms of the second research question, we assumed that the intervention group performs better after the intervention in the credibility evaluation task than the control group. As previous research has shown that adolescents' reading skills contribute to their credibility evaluations (Forzani, 2018; Kanniainen, Kiili, Tolvanen, Aro, & Leppänen, 2019; Kiili et al., 2018), reading fluency and reading comprehension were controlled for in the present study. In addition, recent studies based on PISA-data have revealed that in Finland the gender differences in literacy skills, favoring girls, are especially large (Brozo et al., 2014; Harju-Luukkainen, Vettenranta, Ouakrim-Soivio, & Bernelius, 2016). Further, the study by Forzani (2018) found that girls outperformed boys also in the evaluation of online information. Therefore, gender was also controlled for. Furthermore, it has been shown that the topic of the task plays a role in the evaluation of texts (Bråten et al., 2018a), and for that reason, we also controlled for the topic order.

Recent studies have shown that sourcing in written products is challenging for students at different ages (e.g., Kiili et al., 2020a; Pérez et al., 2018). In terms of the third research question, we therefore assumed that students' credibility evaluations will only rarely be reflected in their written products. Accordingly, we assumed that our teacher-led intervention might not help students to use justifications for credibility in their writings (cf. Pérez et al., 2018).

2. Method

This study reports part of an intervention study in which students were taught online inquiry skills (searching, evaluating and synthesizing online information) within two disciplines, *Social science* and *Science*. Because meanings are constructed in somewhat different ways in different disciplines (Goldman et al., 2016; Shanahan & Shanahan, 2012), students' online inquiry skills were evaluated with separate tasks for each discipline. This study focuses on the *Social science* discipline.[1] In addition, in order to enable the depth of the analyses and careful consideration of instructional implications, this study concentrates on examining the efficacy of the intervention on students' credibility evaluations. Furthermore, this study explores how students' credibility evaluations were reflected in their written products.

2.1. Participants

Sixth graders ($N = 364$) were recruited from ten primary schools (15 classes) in three suburban areas in Finland. Parental permission was received from 345 students (94.78% of the recruited students). Two

[1] Because Science Web pages contained considerably less source features than Social science Web pages, we excluded Science topics from this study.

students were absent from all the tests in this study and one student was excluded for another reason. Hence, the final number of participants was 342 (165 girls and 177 boys). Participants' mean age was 12.13 years ($SD = 0.41$).

2.2. Research design

A quasi-experimental design with pre- and post-tests was applied (Fig. 1). The participating classes were divided into an intervention group and a control group. For practical reasons, convenience sampling was used: intervention group teachers were recruited based on teachers' opportunities and willingness to implement the intervention program. However, control group teachers were not offered a chance to participate in the intervention and they were recruited after the intervention group teachers. Most of them were from different schools than the intervention group teachers. As a result, the intervention group comprised 192 students (90 girls and 102 boys) in eight classes and the control group 150 students (75 girls and 75 boys) in seven classes.

Fig. 1 shows the reading tests and the pre- and post-tests on the online inquiry task that students completed during the intervention study and how the topic order (Topics 1 and 2 in both disciplines) and the test order (Social science and Science) were counterbalanced. The reading tests were conducted one week before the first pre-test. Between the pre- and post-tests, the intervention group participated in a six-week teacher-led intervention program on online inquiry skills while the control group received business-as-usual teaching which follows the Finnish curriculum for basic education. The new curriculum includes broader competencies such as multiliteracies crossing all learning in schools (see Finnish National Board of Education, 2016). In Finland, teachers have a lot of autonomy to realize the aims of the curriculum and they can actualize the curriculum according to their own pedagogical views and their strengths as teachers. The teachers of the control group were not present at the introduction sessions provided for the intervention group teachers, and they received intervention materials after completion of the study. Thus, during the study the control group was not exposed to any of the teaching materials used in the intervention group.

2.3. Measures

2.3.1. Measure of students' credibility evaluation skills

In the pre- and post-tests, students' credibility evaluation skills were assessed as a part of the computer-based online inquiry assessment (González-Ibáñez, Gacitúa, Sormunen, & Kiili, 2017; Sormunen et al., 2018) that consisted of four phases: searching information, evaluating credibility of information, identifying main ideas from single texts, and composing a written product. In the Social science task, the students were asked to explore either *Computer Gaming* (Topic 1) or *Reading on Screen* (Topic 2) in order to compose a newspaper article or an email message (written product) on the advantages and disadvantages of the topic at hand. These topics were chosen owing to their relevance for students' lives and all students were assumed to be somewhat familiar with both.

After receiving the task assignment, students were asked to evaluate their prior knowledge on the topic by answering the 5-point response scale question: *How much knowledge do you think you have about the topic of the given task? (1 = I don't know anything about the topic, 5 = I have much knowledge about the topic)*. There was no statistically significant

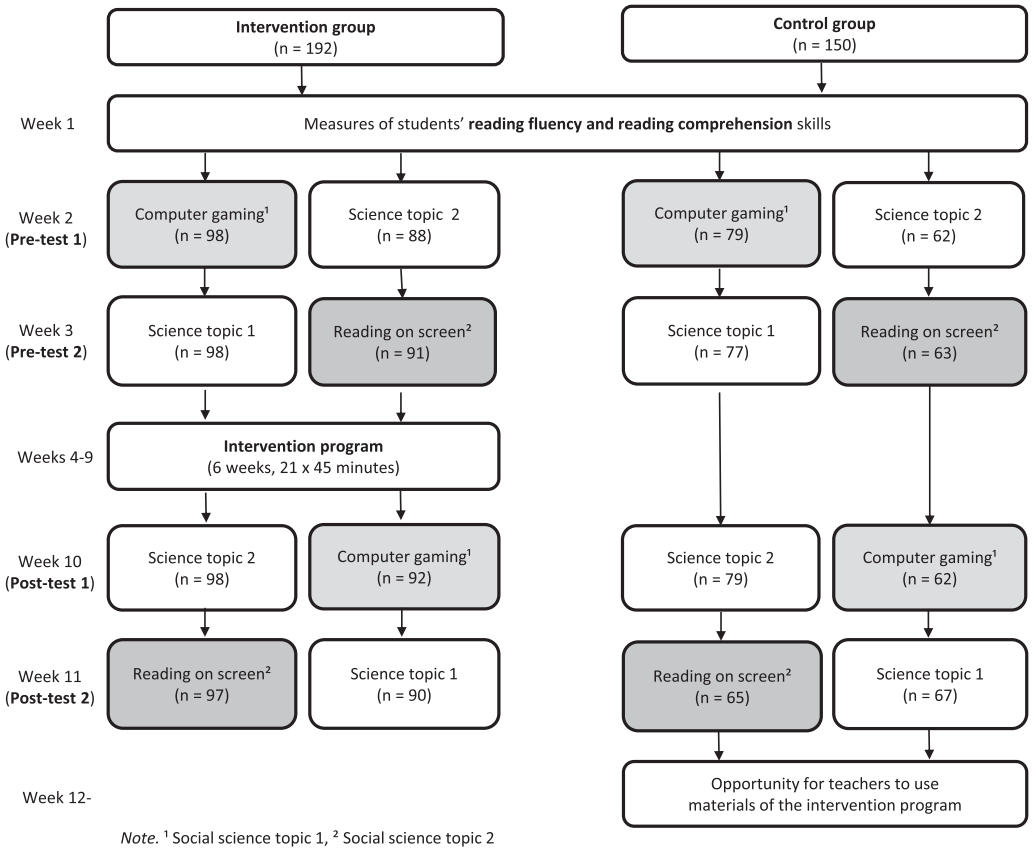


Fig. 1. Research design of the online inquiry study. Pre- and post-tests used in this study are in grey.

difference ($t(304) = -0.742$, $p = .458$) between the self-evaluated prior knowledge on the topics (Computer Gaming, $M = 3.10$, $SD = 0.87$ vs. Reading on Screen, $M = 3.03$, $SD = 0.88$) in the pre-tests.

In the credibility evaluation phase, students evaluated three Web pages designed for the purpose of this study. The author, document type, perspective and position on the topic of each Web page are described in Table 1. First, students were asked to rate the credibility of each Web page on a 5-point response scale consisting of 1–5 stars. Second, students were asked to justify their rating by answering the following open question: *Why do you think so?* (see Appendix A). Students were allowed 7 min to evaluate all three of the Web pages they had read in the previous phase of the task. The computer bleeped 3 min before the time elapsed.

Depending on the test topic, students completed the pre-test either one or two weeks after completing the reading fluency and reading comprehension tests (see Fig. 1). Similarly, the post-test was completed either one or two weeks after the intervention program. When completing the pre- and post-tests, students worked in a classroom with computers. One or two researchers were present in the classroom and observed the students' actions. If students encountered technical problems, the researcher(s) helped them.

2.3.2. Reading measures

To control for students' reading skills, the students' reading fluency and reading comprehension skills were measured before the pre-tests. In the reading test session (45–60 min), the students completed reading fluency and reading comprehension tests in the classroom. The researchers gave the instructions and answered students' practical questions.

Reading fluency was measured with a word chain test (Holopainen, Kairaluoma, Nevala, Ahonen, & Aro, 2004). The word chain test comprised 25 chains, each containing four words written with no spaces in between. The students were asked to separate as many chains into primary words as possible within 90 s. The total score was the number of correctly separated words. The test-retest reliability coefficient for the original test has varied between 0.70 and 0.84 (Holopainen et al., 2004). The total score (0–100) was used in the statistical analyses.

In the *reading comprehension* test (Kajamies, 2017), the students read a text on the diversity of nature and answered three open-ended questions on the main ideas presented in it. Students could consult the text while answering the questions. Students were allowed 15 min to read the text and another 15 min to answer the questions. When needed, students were given 5 min' extra time to finish the test. Students earned 0 to 6 points from each open-ended question, and thus the maximum score was 18 points. The correlations between questions varied from 0.30 to 0.37. The total score of the test (0–18) was used in the statistical analyses. To establish inter-rater reliability, two independent researchers scored 20% of the students' answers ($n = 68$). The following Kappa values (Cohen, 1960) were obtained: 0.90 (Question 1), 0.68 (Question 2), and 0.95 (Question 3).

2.4. Intervention program

The intervention program for online inquiry contained three modules. Module 1 (9 × 45 min lessons) consisted of explicit teaching on searching information, evaluating the credibility of information, and synthesizing information from multiple sources. Each of these component skills was taught in three lessons that followed five phases: 1) *modeling* effective strategies by showing a video in which two virtual students talked aloud while completing an inquiry sub-task, 2) *analyzing* the strategies modeled in the video with a worksheet and sharing thoughts with other students, 3) *discussing* the strategies in a teacher-led session, 4) *practicing* the strategies and, 5) *reflecting on* what one had learned. The first three phases were conducted without computers to keep the students' attention on the main points of the strategies. In the lessons regarding searching for information, the students analyzed search queries differing in their quality. Further, the lessons regarding synthesizing information from multiple sources focused on integration of ideas from multiple Web pages. The students were taught how to compose fluent texts and use connecting words to integrate ideas. This study focused on credibility evaluation lessons that are described in more detail below.

In Module 1, three 45 min scripted lessons concerning credibility evaluation were implemented in two sessions (one 45 min lesson and one 2 × 45 min lessons). In designing the lessons, we employed contrasting cases and materials (cf. Braasch, Bråten, Strømsø, Anmarkrud, & Ferguson, 2013; Bråten et al., 2019). In addition, all materials covered source-based and content-based aspects of credibility evaluation at a level appropriate for sixth graders.

In the first 45 min lesson, students watched a video where two virtual students modeled evaluation strategies by thinking aloud when evaluating the Web news article. The virtual students varied in the versatility and sophistication of the evaluation strategies they used. The skillful virtual student employed sophisticated strategies, such as quality of evidence on the Web page, and date and type of the Web page. The other, not that skilled virtual student, used more superficial strategies, such as referring to the amount of the text on the Web page and appearance of the Web page. While watching the video, students were prompted to analyze and compare virtual students' evaluation strategies with the worksheet followed by the discussions of the strategies with their partners and the teacher (see Table 2).

In the 90-min double lesson, we used two contrasting Web pages on health effects of energy drinks that differed in their credibility. The more credible page was written by a researcher who responded to FAQs at the University website. The less credible Web page was a commercial press release written by a head of marketing (see Kiili et al., 2018). While

Table 1
Features of the Web pages evaluated by the students.

Topic	Title	Authors	Type	Perspective	Position on the Topic
Computer Gaming	Computer gaming has both advantages and disadvantages for health	Pediatrician	Blog by an expert	Health	For and against
	Learning games researchers met at the University of Tampere	University researcher responsible for organizing the event	Press release	Learning	For
	Violent computer games increase hostile behaviour	Journalist	News	Violence	Against
Reading on Screen	Digital book is easy to carry but digital texts are not easily remembered	Journalist	News	Learning	For and against
	Every student should have a chance to read digitally	School principal	Opinion piece in a newspaper	Equality	For
	Excessive screen time is hard on the eyes	Health research center	Official page of an organization	Health	Against

Table 2
Module 1: Credibility evaluation (3 × 45 min lessons).

Lesson	Instructional methods	Materials	Activities	Prompts for students or guiding questions for teachers
Lesson 1	Modeling strategies Analyzing modeled strategies Discussing strategies	Web news article [1]: "New Study: Later school mornings have significant influence on students' Video where two virtual students, Tiina and Toni, evaluated the credibility of the Web news article. The students varied in how versatile and sophisticated evaluation strategies they used. Worksheet [1]	a. Students read the news article and considered its credibility. b. Students watched the video that modeled strategies and analyzed these strategies with the worksheet. c. Students discussed their answers with a partner/partners. d. A teacher-led discussion on the evaluation strategies and the content of the article.	Prompts in the worksheet: 1. What evaluation strategies the virtual students used in the video? (identification from the list) 2. Compare how Tiina and Toni evaluated the news article. Which of them did it more thoroughly? Why do you think so? 3. What new ways to evaluate Web pages' credibility you learned?
Lessons 2-3	Practicing strategies Discussing strategies Reflecting learning	Two Web pages [1] concerning energy drinks and healthy effects varying with their quality (an academic page and a commercial page) [3] Worksheets [1 or 2] Learning diary [1]	a. Students read two Web pages and filled in the worksheets. b. A teacher-led discussion of the credibility of evaluated Web pages. c. Students individually evaluated their learning in their diaries.	Prompts in the worksheets: 1. Who is the author/publisher of the Web page? 2. Is she/he an expert or not? Why do you think so? 3. What is the purpose of the Web page? (multiple-choice question) 4. How do you know whether the author's message seems reliable or not reliable? 5. What do you think about the information in the Web page? (multi-choice question about use of evidence) 6. How credible is the Web page? (rating) Why do you think so? Guiding questions in the teacher manual: - How did the Web pages differ? - How did they talk about the energy drinks? - Why is it important to compare information from different Web pages? - Why is it important to evaluate Web pages in different ways? - How can you assure which Web page contains information that is more credible? Prompts in the students' diaries: - questions fostering self-evaluation of learning and reflection of learning experiences

Note: [1] on paper, [2] digital, [3] Web pages were designed for assessment purposes (Kiili et al., 2018).

reading the Web pages, students were prompted to evaluate author expertise, purpose of the text and use of evidence with the worksheet. After that, the credibility of the Web pages was compared in a teacher-led discussion (see Table 2).

The taught credibility evaluation strategies were applied and practiced in Modules 2 and 3. In Module 2 (4 × 45 min lessons), the students practiced the online inquiry skills in a social science project on the advantages and disadvantages of social media. In the project, students 1) activated their prior knowledge on social media, 2) searched for information with Google Custom Search Engine, which contained a limited number of authentic Web pages, 3) evaluated the credibility and noted the advantages and disadvantages of social media from their reading of the selected Web pages and, 5) engaged in teacher-led discussion in which different perspectives on social media were compared and contrasted.

In Module 3 (8 × 45 min lessons), the students practiced the taught online inquiry skills in a science project on energy. The project was longer than the social science project (Module 2) as it was more closely adapted to the curriculum. The project followed the same phases as the social science project but with some differences in implementation: the students 1) activated their prior knowledge on energy, 2) searched for information in open Web sources, 3) evaluated the credibility of selected Web pages, and 4) wrote and presented syntheses based on multiple Web pages. During both projects, the students worked individually and in small groups, and their work was supported with worksheets.

Fidelity to the intervention program was assured in several ways (see McKenna, Flower, & Ciullo, 2014). The teachers of the intervention group (N = 9) received a detailed intervention manual that they were asked to follow. The manual included a short theoretical introduction to the components of online inquiry, the goals and phases of the lessons, and instructional materials and methods. We decided to model the strategies using videos instead of teacher modeling to ensure that all students received exactly the same examples of the strategies. Before each module, the teachers were given a short induction session (30–60 min) on the materials, tasks and teaching methods. The teachers were also assigned a researcher to contact if they had any further questions about the lessons.

We observed five teachers' lessons on credibility evaluation in Module 1. In addition, other lessons from Modules 1 to 3 were followed, making a total of 44 observations of lessons. The other four teachers were interviewed after each module. During the intervention program, the teachers recorded in a diary any deviations from the plan and their observations of the efficacy and adequacy of the tasks, materials and activities. All the students' worksheets were also collected after each Module. The observations, interviews, diaries and worksheets revealed that the intervention program was mostly conducted as planned. The teachers felt that some tasks in the worksheets were too difficult or complex for the students (in Modules 2 and 3) and therefore they made some pedagogical modifications to the plan.

2.5. Data analyses

2.5.1. Students' justifications for their credibility ratings

Students' justifications for their credibility ratings from the pre- and post-tests were identified, categorized, and counted. The unit of analysis was an expression containing a justification. The identified justifications were divided into four categories. Two categories (Expertise of the Source and Other Source Features) concerned the evaluation of the *Source Features* (e.g., Bråten et al., 2018a; Britt & Aglinskas, 2002; Britt et al., 2014) and two (Argumentation in the Text and Other Aspects of the

Content) the evaluation of the *Quality of Content* (e.g., Braasch et al., 2013; Britt et al., 2014; Judd, Farrow, & Tims, 2006; Metzger, 2007) (Table 3).

After categorization, the number of the justifications in each of the four categories (Expertise of the Source, Other Source Features, Argumentation in the Text, and Other Aspects of the Content) was counted. The students received one point for each relevant justification that was in accordance with their credibility rating. Four count variables were formed for both the pre- and post-tests, respectively (8 count variables in total). These count variables were constructed based on the total number of relevant justifications across the three evaluated Web pages. In statistical analyses, the post-test count variables were used as dependent variables and the pre-test count variables used as control variables.

The reliability of the categorization was calculated for a random sample of 15% of the justifications. The first rater identified and categorized the justifications in the students' answers and the other rater categorized the justifications identified by the first rater. The inter-rater agreement for the categorization was 0.90 (Cohen's kappa; Cohen, 1960). Disagreements were resolved through discussion between the raters.

2.5.2. Students' use of their credibility evaluations in written products

To explore how students' credibility evaluations were reflected in their written products, we examined whether students' justification for their credibility rating or a part of it appeared in their writings. If we found an overlap, it was categorized as representing one of the main categories of the justifications for credibility ratings: Source Features or Quality of Content. The examples below illustrate how students used their justifications in their essays.

Example 1 (Source Features).

Student's justification of credibility rating:

This is an opinion piece but it is written by a principal. (ID 2007).

Use of justification in the written product:

The principal Ulla-Maija Lehola from Comprehensive School of Helsinki says that reading on screen should be taught for students at school in order to give them sufficient skills to use Internet and different media in their learning. (ID 2007).

Example 2 (Quality of Content).

Student's justification of credibility rating:

The text gives a wide understanding of advantages and disadvantages of computer gaming. (ID 3354).

Use of justification in the written product:

Some people think that sitting by the computer is harmful for your health whereas some people think that computer gaming can develop your skills. Both perspectives are correct. (ID 3354).

The effect of the intervention on the association between students' credibility evaluations and their written products was examined with non-parametric Wilcoxon's test separately for the intervention and control groups. In addition, Cohen's *d* and its 95% confidence intervals were computed for both groups.

2.5.3. Statistical analyses

The effect of the intervention on the students' credibility evaluation skills was examined by a negative binomial regression analysis (Coxe, West, & Aiken, 2009; Gardner, Mulvey, & Shaw, 1995). It is a suitable method here, as the four dependent variables (i.e., the number of justifications in the four post-test categories: Expertise of the Source, Other Source Features, Argumentation in the Text, and Other Aspects of the Content) were by their nature non-normally distributed count variables. These variables also showed large over-dispersion, meaning that the variance of each dependent variable was larger than its mean.

The dependent variables were analyzed separately. In each analysis, the corresponding Pre-test score (number of justifications) was controlled for. In all four analyses, Group (intervention or control) was used as the independent variable, whereas Gender, Reading fluency, Reading comprehension, Topic order, and Test order in the pre-test were

Table 3
Categories of students' justifications.

Main category	Sub-category	Description	Example	Interpretation
Source Features	Expertise of the Source	Student justifies his or her credibility rating with authors'/publishers' credentials, affiliation, experience or expertise.	*I think this Web page is credible because it includes the ideas of a pediatrician.* (Student 279, topic: Computer Gaming) *The Web page has been created by a health research center.* (Student 292, topic: Reading on Screen)	Being a pediatrician is a credential of the author/publisher (page is a blog by an expert). The health research center is an affiliation of the author and publisher.
	Other Source Features	Student justifies his or her credibility rating with the date, appearance or type of the Web page, or availability of contact information or references.	*The Web page includes new information.* (Student 10, topic: Computer Gaming) *The Web page seems to be a rather credible news page.* (Student 197, topic: Reading on Screen)	New information refers to the date of the Web page (page was published in 2016). The Web page type is news in an online newspaper.
Quality of Content	Argumentation in the Text	Student justifies his or her credibility rating on research basis, quality of evidence, consideration of both sides (negative and positive) of an issue or argumentative purpose of the text on the Web page.	*It tells about the negative effects of games but also recommends a useful game.* (Student 212, topic: Computer Gaming) *Altogether 1492 adolescents participated in that study lasting over 4 years.* (Student 206, topic: Reading on Screen)	Student refers to the consideration of both sides (negative and positive) of an issue on the Web page. Student refers to the research-basis and quality of evidence on the Web page.
	Other Aspects of the Content	Student justifies his or her credibility rating with correspondence with his/her own experiences or prior knowledge, or amount of the text or writing style on the Web page.	*Gaming affects one's fitness and health, I have noticed it also myself.* (Student 103, topic: Computer Gaming) *I think this Web page is quite credible because its' style of writing is correct.* (Student 301, topic: Reading on Screen)	Student refers to the correspondence with her own experiences. Student refers to the writing style on the Web page.

controlled for. The descriptive statistics of all the employed variables are presented in Table 4.

Data were hierarchical in nature: that is, the students were nested within classes. Intra-class correlations (Heck, 2001; Muthén, 1991) in the pre-tests (Expertise of the Source = 0.07, Other Source Features = 0.17, Argumentation in the Text = 0.02, Other Aspects of the Content = 0.06) suggested the presence of variation between the classes in the justification scores, particularly for the variable Other Source Features. Therefore, multilevel modeling (Muthén, 1997) was used with class as a clustering variable. The variation between the classes was taken into account by estimating the means of the dependent variables at the class level (i.e., between-level) as random. The actual negative binomial regression was conducted as a student-level (i.e., within-level) analysis.

Negative binomial regression analysis models the log of the expected count of justifications in each post-test category (dependent variables) as a function of the independent/control variables (Coxe et al., 2009; Gardner et al., 1995). For ease of interpretation, the regression coefficients were presented as incident rate ratios (IRRs) which were obtained by exponentiating the regression coefficients using base *e*. For example, for a dichotomous independent variable (i.e., Group), the IRR represents the change in the expected rate of justifications in a specific post-test category when the value of the independent variable changes from 0 to 1. An IRR greater than 1 indicates how many times greater the expected rate of justifications in the post-test category is for students with an independent variable value of 1 (i.e., Group: intervention) than for those with the value 0 (i.e., Group: control).

In contrast, an IRR smaller than 1 indicates that the expected rate of justifications in the post-test category is greater for those participants with an independent variable value of 0 (i.e., Group: control) than those with the value 1 (i.e., Group: intervention). With continuous control variables (e.g., Reading fluency score), the IRR represents the change in the expected rate of justifications in a post-test category when the value of the control variable (e.g., Reading fluency score) increases by one unit. The statistical significance of the IRRs was determined by computing 95% confidence intervals for each IRR. An IRR differs statistically significantly from the value 1 if its confidence interval does not include the value 1.

All statistical analyses were conducted using the Mplus statistical package (version 8.0; Muthén & Muthén, 1998–2017) with Full--Information-Maximum-Likelihood (FIML) procedure (Enders, 2010). FIML uses all available information in the data to estimate the model without imputing missing values. Model parameters were estimated using maximum likelihood estimation with non-normality robust standard errors (MLR) (Muthén & Muthén, 1998–2017).

3. Results

3.1. Students' performance on the credibility evaluation task

Students' performance in the reading tests and pre- and post-tests is presented in Tables 4 and 5. In the pre-tests, the highest scores were for the category Expertise of the Source, suggesting that students most often justified their credibility ratings by reference to the expertise of the source (Table 4, variable 5). The variation in the number of justifications was also largest in this category, indicating that some students found many justifications related to this category and others none. In contrast, students seemed to find evaluation of the other credibility aspects (Other Source Features, Argumentation in the Text and Other Aspects of the Content) rather challenging (Table 4, variables 6–8). The control group outperformed the intervention group only in the category of Other Source Features (Table 5). In all the other pre-test categories and in the reading fluency and reading comprehension tests, the intervention and the control groups performed similarly, indicating no group differences at baseline.

Table 4
Descriptive statistics for the variables and their correlations ($N = 316$–342).

Variables (observed range)	Mean (*SD*)/%	2.	3.	4.	5.	6.	7.	8.	9.	10.	11.	12.	13.	14.
Dependent variables														
1. Post-test: Expertise of the Source (0–9)	1.84 (2.05)	.23**	.18**	-.17**	.51**	.28**	.18**	-.14*	.32**	.30**	-.01	-.10	-.01	.09
2. Post-test: Other Source Features (0–6)	0.56 (0.95)		.04	-.05	.15*	.34**	.06	-.05	.19**	.15**	-.07	-.16**	-.05	.09
3. Post-test: Argumentation in the text (0–4)	0.29 (0.64)			-.02	.10	.10	.11	-.04	.19**	.20**	-.24**	-.09	-.24**	-.01
4. Post-test: Other aspects of the content (0–3)	0.39 (0.72)				-.12*	.03	-.02	.32**	.01	.02	-.01	-.03	-.01	-.03
Control variables														
5. Pre-test: Expertise of the Source (0–10)	1.56 (1.95)					.27**	.21**	-.23**	.34**	.16**	.21**	-.11*	.22**	-.13*
6. Pre-test: Other Source Features (0–8)	0.52 (1.05)						.03	-.07	.25**	.13*	.11	-.11*	.13*	-.18*
7. Pre-test: Argumentation in the text (0–4)	0.33 (0.71)							-.07	.16**	.16**	.20**	-.11	.16**	-.07
8. Pre-test: Other aspects of the content (0–5)	0.39 (0.79)								-.08	.11*	.10	.01	.07	-.02
9. Reading fluency (10–89)	38.32 (14.00)									.40**	-.00	-.33**	.03	-.07
10. Reading comprehension (0–16)	8.81 (3.62)										-.09	-.22**	-.09	.04
11. Topic order (screen - gaming/gaming - screen)	48.37/51.63											-.09	.88**	-.00
12. Gender (female/male)	48.25/51.75												-.06	.03
13. Test order in pre-test (first/second)	53.60/46.40													-.03
Independent variable														
14. Group (intervention/control)	56.14/43.86													

Note: * = $p < .05$; ** = $p < .01$; *** = $p < .001$.

3.2. Effect of an intervention on the students' performance on the evaluation task

After controlling for the background and pre-test variables, the results showed that the intervention helped students to better justify their credibility ratings with reference to Expertise of the Source and Other Source Features (Table 6). The intervention group justified their credibility ratings 1.52 times more often with expertise of the source and 1.83 times more often with other source features than controls.[2] However, the intervention group and control group did not differ in the numbers of justifications for their credibility ratings related to argumentation in the text or other aspects of the content.

Of the control variables (Table 6), the pre-test scores (number of justifications) explained the post-test scores; thus, the more justifications a student had in the pre-test, the more she/he also had in the post-test in each corresponding category. Furthermore, the students' reading comprehension skills were associated with the number of justifications in the post-test in the categories Expertise of the Source and Argumentation in the Text: the better reading comprehension skills a student had, the more often he/she justified his/her credibility ratings with reference to expertise of the source and the argumentation in the text.

Topic order was associated with only one of the post-test scores. The students who completed the Reading on Screen task first scored better in justifying credibility by reference to argumentation in the text than the students who completed the Computer Gaming task first. The other control variables (i.e., Reading fluency, Gender, and Test order in the pre-test) were not associated with the number of justifications in any of the four justification categories in the post-test (Table 6).

3.3. Students' use of their credibility evaluations in written products

Table 7 presents the results of students' use of justifications for credibility ratings in their written products between intervention and control groups and results of the Wilcoxon's tests computed separately for the groups. In the pre- and post-tests, students' justifications were rarely reflected in their writings. Of all students, 8.4% in the pre-tests and 10.8% in the post-tests utilized justifications related to Source Features or Quality of Content in their essays. Wilcoxon's tests for change in Source Features and Quality of Content did not show statistically significant change either in the intervention or control group. In addition, the confidence intervals of the effect sizes (Cohen's *d*) included 0 in both groups, thus showing that the effect of change in both groups was 0. These confidence intervals also overlapped in both categories (Source Features and Quality of Content). Therefore, it can be concluded that the intervention did not increase students' use of justifications in their written products (Table 7).

4. Discussion

The present study, with 340 students, evaluated the efficacy of a teacher-led intervention program that combined different instructional methods, such as modeling, prompting, and discussing evaluation strategies, with the aim of improving students' credibility evaluation skills. We first discuss the main findings and then evaluate the instructional design and implementation of the intervention program. Finally, we consider the limitations of this study and discuss the instructional implications of the findings.

Our intervention program improved students' skills to evaluate source features when considering the credibility Web pages. After participating in the intervention program, students more often justified their credibility ratings with reference to source features. After ruling out plausible alternative data-related explanations for the effects (i.e., Pre-test scores, Reading skills, Gender, Topic order and Test order in the pre-test), the intervention group students presented 1.52 times more justifications related to expertise of the source and 1.83 more justifications related to other source features than controls. Bråten et al. (2018a) emphasize that the ability to use source features in credibility evaluation is particularly important when students use the Internet to explore controversial issues about which they have little prior knowledge. Without relevant prior knowledge, it is very difficult to evaluate, e.g., the accuracy of information, which suggests that readers are left to rely on source features when evaluating the credibility of Web pages.

However, we observed no improvement in students' content-based evaluation of credibility. There might be several reasons for this finding. First, paying attention to source features, i.e., author and her/his credentials, publisher and date (Britt et al., 2014; Perfetti et al., 1999), can be more concretely modeled and taught compared to content-based evaluation strategies. To be effective, teaching credibility evaluation of argumentation would have required, for example, in-depth knowledge of argument structure and what counts as high quality evidence and why. Second, students can quite easily find source features in Web pages whereas evaluation of argumentation requires careful reading of the texts and identification of claims and related evidence (Britt et al., 2014). In addition, the evaluation of ideas in light of one's prior knowledge or experience requires thoughtful and reflective reading.

Our result on the growth in the use of source features when justifying credibility was favorable and encouraging, as attention to and identification of source features is a prerequisite for other sourcing activities, such as using source features to interpret content (Brante & Strømsø, 2018). The use of source features in interpreting content is, however, a very demanding task for young students (Britt et al., 2014; Macedo-Rouet et al., 2013) and requires teaching that systematically builds on the lower-level skills, i.e., attending to source features, already acquired.

Despite the positive finding in the increase in justifications related to source features in their assessments of credibility, the proportion of students whose performance remained low was quite high (see also Kingsley et al., 2015; Zhang & Duke, 2011). Our results further showed that reading comprehension skills predicted students' performance in their evaluation of the expertise of the source and the argumentation in the text. Hence, it is possible that poor reading skills also hinder students' performance in their justifications for the credibility of Web pages, as suggested in a recent study by Kanniainen et al. (2019).

In the present study, we also examined how students' justifications for their credibility ratings were reflected in the written products. As assumed, students very rarely used evaluations of sources and content in their writings. This is in accordance with previous studies showing that, in general, citing or evaluating sources in the written products is not that common practice, in particular, among younger students (Kiili et al., 2020a; Pérez et al., 2018). In terms of synthesizing information from multiple Web pages, the focus of our intervention was not on teaching how to utilize credibility evaluations in writing but on integration of ideas with connecting words. Accordingly, our intervention did not increase students' use of justifications for credibility evaluations in their written products. As the use of credibility evaluations in the written products was rare both in the pre- and post-tests, it is evident that students need more explicit support to understand the connections between evaluations of Web pages and writing from

[2] To more thoroughly examine the possibility that the intervention effect could have resulted from the lower scores of the control group in Expertise of the Source and Other Source Features, Cohen's *d* for repeated measures was computed separately for the intervention and control groups (Morris & DeShon, 2002). The *d* for Expertise of the Source was 0.35 (95% confidence interval: [0.14; 0.55]) for the intervention group and -0.12 (95% confidence interval: [-0.37; 0.13]) for the control group. The *d* for Other Source Features was 0.32 (95% confidence interval: [0.11; 0.52]) for the intervention group and -0.22 (95% confidence interval: [-0.47; 0.03]) for the control group. As the confidence intervals of the intervention and control groups in either variable do not overlap, it can be concluded that the effect of the intervention group (in Expertise of the Source and Other Source Features) was greater than the corresponding effect of the control group.

Table 5
Pre-test, post-test and reading test scores in the intervention and control group.

	Intervention group (n = 189–191)			Control group (n = 127–143)			Test result
	M	*SD*	*Md*	*M*	*SD*	*Md*	
Pre-test measures							
Expertise of the Source	1.35	1.75	0	1.85	2.17	1	U = 11,936.50, p = .066, r = −.10
Other Source Features	0.35	0.65	0	0.74	1.40	0	U = 12,041.00, p = .048, r = −.11
Argumentation in the Text	0.29	0.66	0	0.39	0.78	0	U = 12,842.50, p = .355, r = −.05
Other Aspects of the Content	0.38	0.81	0	0.40	0.77	0	U = 13,002.50, p = .522, r = −.04
Reading test measures							
Reading fluency	37.48	13.52		39.45	14.57		t(332) = 1.28, p = .203, d = 0.14
Reading comprehension	8.93	3.73		8.65	3.47		t(331) = −0.71, p = .480, d = 0.08
Post-test measures							
Expertise of the Source	1.98	2.10	2	1.62	1.98	1	
Other Source Features	0.63	0.93	0	0.46	0.97	0	
Argumentation in the Text	0.29	0.66	0	0.30	0.61	0	
Other Aspects of the Content	0.37	0.68	0	0.41	0.77	0	

Note. The effect of an intervention cannot be calculated based on this table as the background variables are not controlled for.

Table 6
Results of the multilevel negative binomial regression analysis for the association between control variables, independent variable (group) and post-test measures.

	Post-test measures			
	Expertise of the Source	Other Source Features	Argumentation in the Text	Other Aspects of the Content
	IRR*95% *CI*	IRR* 95% *CI*	IRR* 95% *CI*	IRR* 95% *CI*
Pre-test	**1.28**			
Expertise of the Source	**[1.19; 1.36]**			
Pre-test:		**1.39**		
Other Source Features		**[1.25; 1.55]**		
Pre-test:			**1.35**	
Argumentation in the Text			**[1.05; 1.74]**	
Pre-test:				**1.74**
Other Aspects of the Content				**[1.37; 2.20]**
Reading fluency	1.01	1.01	1.02	1.01
	[1.00; 1.02]	[0.99; 1.02]	[1.00; 1.04]	[0.98; 1.03]
Reading comprehension	**1.06**	1.02	**1.09**	0.98
	[1.03; 1.10]	[0.97; 1.07]	**[1.02; 1.15]**	[0.92; 1.04]
Gender	1.00	0.69	0.87	0.92
(0 = female, 1 = male)	[0.77; 1.30]	[0.45; 1.05]	[0.55; 1.38]	[0.66; 1.29]
Topic order	1.01	0.65	**0.43**	0.74
(0 = screen-gaming, 1 = gaming-screen)	[0.51; 2.01]	[0.40; 1.06]	**[0.20; 0.90]**	[0.25; 2.21]
Test order in the pre-test	0.85	1.16	0.65	1.21
(0 = second, 1 = first)	[0.45; 1.61]	[0.73; 1.82]	[0.31; 1.35]	[0.50; 2.93]
Group	**1.52**	**1.83**	1.01	1.00
(0 = control group, 1 = intervention)	**[1.15; 2.00]**	**[1.29; 2.60]**	[0.67; 1.53]	[0.57; 1.74]

* IRR = incident rate ratio (IRR differs statistically significantly from value 1 if its confidence interval does not include value 1).

multiple pages.

4.1. Evaluation of the intervention

In this section, to contribute further to research in the field, we evaluate the intervention program in terms of its instructional design and implementation. Finally, we evaluate the measure used to assess the efficacy of the program.

In this study, credibility evaluation of Web pages was taught as a part of an intervention program designed to support students' online inquiry skills. This was done to provide an authentic context for the evaluation activities. After explicit teaching of different online inquiry skills, students were able to apply their new skills in two inquiry projects. This meant that the 21-lesson intervention program, which included instruction in several different kinds of skills (searching, evaluating, synthesizing) was rather long. This might have been experienced as overwhelming, especially by students who find online inquiry a struggle. Furthermore, the explicit teaching of evaluation skills was accomplished in only three lessons, which might not be sufficient time for all students to learn a skill as complex as credibility evaluation (cf. Argelagós & Pifarré, 2012; Walraven et al., 2010).

Moreover, the credibility evaluation lessons were combined with different instructional methods, such as modeling, prompting, and discussing strategies. The aim of combining these was to highlight different aspects of the evaluation of the credibility of Web pages and to deepen students' understanding of critical evaluation. To model the evaluation strategies (Coiro, 2011; Davey, 1983), we provided a video where two virtual students performed a credibility evaluation task. A modeling video might be especially useful for teachers who do not feel comfortable modeling strategies themselves. A video can also be a motivating tool for students (Choi & Johnson, 2005). However, the video modeled multiple evaluation strategies, which may have caused some students cognitive overload and thereby hindered their learning. On the other hand, students were able to watch the video multiple times.

In the lessons, prompts on the worksheets were used to direct students' attention to different aspects of credibility in performing their given tasks during the intervention. To maximize the benefit gained from prompts, students' responses need to be discussed in the classroom (Macedo-Rouet et al., 2019; Pérez et al., 2018). However, according to our observations, the discussions remained rather shallow and teachers

Table 7
Students' use of justifications for credibility in their written products.

	Category of justifications used in the written product	Pre-test			Post-test			Wilcoxon test result (Z) Effect size (d)
		M	*SD*	*Md*	*M*	*SD*	*Md*	
Intervention group ($n = 189$)	Source Features	0.04	0.22	0	0.08	0.29	0	$Z = 1.56$, $p = .120$ $d = 0.13$, 95% CI = [-0.08; 0.33]
	Quality of Content	0.05	0.29	0	0.07	0.28	0	$Z = 0.97$, $p = .333$ $d = 0.07$, 95% CI = [-0.13: 0.28]
Control group ($n = 143$)	Source Features	0.06	0.23	0	0.02	0.13	0	$Z = -1.67$, $p = .096$ $d = -0.12$, 95% CI = [-0.37; 0.13]
	Quality of Content	0.06	0.23	0	0.07	0.26	0	$Z = 0.24$, $p = .808$ $d = 0.00$, 95% CI = [-0.25; 0.25]

did not necessarily take advantage of the guiding questions provided in the intervention manual. Thus, it seems that, for teachers, orchestrating and motivating classroom discussions was one the most challenging features of the intervention (see also Walraven et al., 2010; 2013).

To investigate the efficacy of the intervention, we used a larger online inquiry task during which the students were asked to justify their credibility ratings of three Web pages. The benefit of such a measure is that it reveals the criteria that students apply when evaluating the credibility of Web pages during authentic online inquiry tasks (cf. Brante & Strømsø, 2018). One downside of this approach is that the students might not have perceived that the evaluation task was as important as the final writing task in which they were expected to report their findings. Moreover, some students might have had difficulties expressing their thinking when writing under the pressure of time (see Macedo-Rouet et al., 2019). Future studies could use a repertoire of measures that would enable all students to express themselves. Finally, while previous intervention studies on evaluation at the primary and lower secondary levels have used different types of outcome measures, such as credibility scales for evaluating single Web pages or rank-ordering Web pages according to their credibility (see Brante & Strømsø, 2018), comparing the efficacy of different interventions is challenging.

4.2. Limitations and future research

This study has its limitations, which should be taken into account in future studies. First, the intervention comprised two different domains (Social science and Science), meaning two pre-tests and two post-tests to be completed by every student. Not all the students may have been equally motivated to complete the post-tests. This was probably true of both the intervention and control groups, especially as the mean performance of the control group declined over time (see Table 5). However, the positive effect was bigger for the intervention than control group in Expertise of the Source and Other Source Features.

Second, the long-term effects of the intervention were not measured. A follow-up could have given information on how well maintained the changes were (e.g., Bråten et al., 2019; Pérez et al., 2018). In our study design, the students had already done an online inquiry test four times, and hence may have found the inclusion of two additional delayed post-tests overwhelming. On the other hand, some studies on evaluation skills (Kroustallaki et al., 2015; Walraven et al., 2013) have reported that a transfer effect is hard to achieve. In future studies, both the long-term and transfer effects could be measured.

Finally, students' prior knowledge on the topics (Computer Gaming and Reading on Screen) was only measured with one self-evaluation question indicating no statistically significant difference between the topics in the pre-tests. However, self-evaluation question can not be regarded as an objective measure of prior knowledge. In the statistical analysis topic order was controlled for, and therefore, we were able to account for possible differences in the difficulty level of the topics.

4.3. Instructional implications

The intervention program applied in this study, while promising, could be developed further in several ways. First, although our students found most of the credibility evaluation strategies difficult, they varied in their ability to justify their credibility ratings, especially with reference to source features. Therefore, teachers should tailor their instruction to individual student needs. Low-performing students may need more scaffolding to understand how to evaluate online information. For example, teachers could begin by showing them where to find author information on different types of Web pages (Paul et al., 2017). These students may also benefit from prompts that explicitly guide them to focus on relevant source features (cf. Kammerer et al., 2016). In this study, all the learning materials (e.g., video, worksheets) used in the intervention program mixed source-based and content-based aspects of credibility evaluation. In the future, it might be more effective to concentrate on just a few aspects of evaluation in one lesson (see Kingsley et al., 2015; Pérez et al., 2018). This might help low-performing students to focus more deeply on a limited number of issues at a time. However, the more skillful students could benefit from materials and discussions that demonstrate the interaction of source-based and content-based aspects of credibility evaluation.

Second, more emphasis should be placed on the professional development of teachers who implement intervention programs in classrooms. In order to engage in critical discussions with their students, teachers themselves need to feel comfortable with the various evaluation strategies (cf. Paul et al., 2017). Teachers also need both knowledge-based and pedagogical abilities to adapt their teaching to their students' needs and to react appropriately to students' comments and answers. In our study, the teachers were offered only a short introduction to the intervention manual and learning materials before conducting the lessons. Instead, teachers should be provided with tools not only to help them orchestrate high-quality discussions (Pérez et al., 2018; Walraven et al., 2010; 2013) but also to guide them in giving students feedback on their evaluation skills during lessons (Macedo-Rouet et al., 2019; Paul et al., 2017) and to motivate their students to regularly evaluate the credibility of Web pages (Brante & Strømsø, 2018; Pérez et al., 2018).

CRediT authorship contribution statement

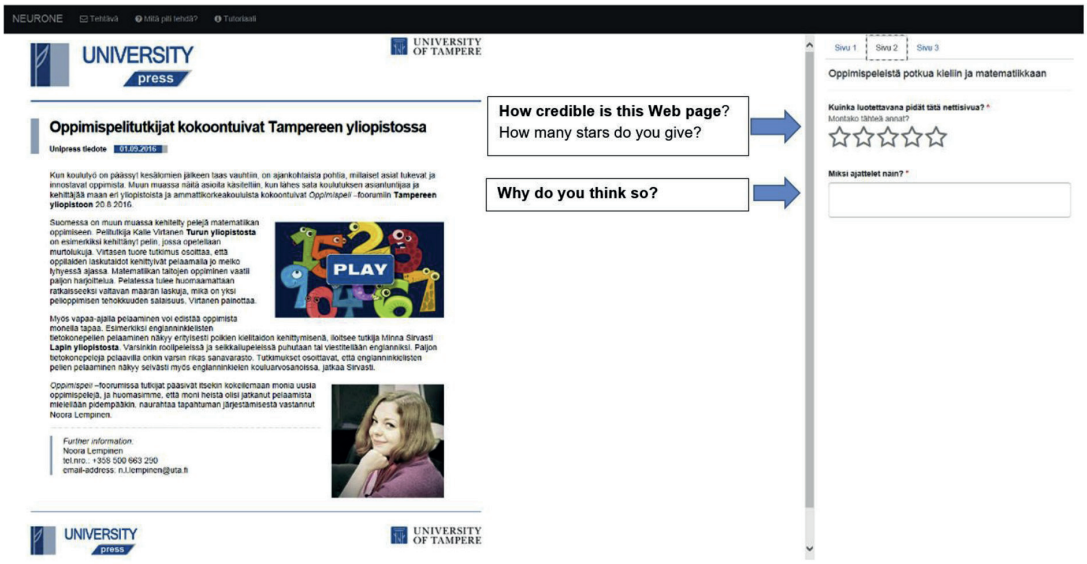
Elina K. Hämäläinen: Conceptualization, Methodology, Formal analysis, Investigation, Resources, Writing - original draft, Writing - review & editing, Visualization. **Carita Kiili:** Conceptualization, Methodology, Resources, Supervision. **Miika Marttunen:** Conceptualization, Methodology, Supervision. **Eija Räikkönen:** Methodology, Formal analysis, Supervision. **Roberto González-Ibáñez:** Software, Data curation. **Paavo H.T. Leppänen:** Project administration, Funding acquisition, Supervision.

Acknowledgements

The Authors thank Michael Freeman for his valuable comments on the language, and the teachers and the students who participated in the study. This research was funded by the Academy of Finland (Project Number: 294197).

Appendix A

Example of the Web page evaluated by the students and the measure of students' credibility evaluation skills.



References

Applebee, A. N., Langer, J. A., Nystrand, M., & Gamoran, A. (2003). Discussion-based approaches to developing understanding: Classroom instruction and student performance in middle and high school English. *American Educational Research Journal, 40*(3), 685–730. https://doi.org/10.3102/00028312040003685.

Argelagós, 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.

Berthold, K., Nückles, M., & Renkl, A. (2007). Do learning protocols support learning strategies and outcomes? The role of cognitive and metacognitive prompts. *Learning and Instruction, 17*(5), 564–577. https://doi.org/10.1016/j.learninstruc.2007.09.007.

Braasch, J. L. G., Bråten, I., Strømsø, H. I., Anmarkrud, Ø., & Ferguson, L. E. (2013). Promoting secondary school students' evaluation of source features of multiple documents. *Contemporary Educational Psychology, 38*(3), 180–195. https://doi.org/10.1016/j.cedpsych.2013.03.003.

Brand-Gruwel, S., & van Strien, J. L. H. (2018). Instruction to promote information problem solving on the Internet in primary and secondary education: A systematic literature review. In J. L. G. Braasch, I. Bråten, & M. T. McCrudden (Eds.), *Handbook of multiple source use* (pp. 401–414). New York, NW: Routledge.

Brand-Gruwel, S., Wopereis, I., & Walraven, A. (2009). A descriptive model of information problem solving while using Internet. *Computers & Education, 53*(3), 1207–1217. https://doi.org/10.1016/j.compedu.2009.06.004.

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. G. (2018). The role of conflict in multiple source use. In J. L. G. Braasch, I. Bråten, & M. T. McCrudden (Eds.), *Handbook of multiple source use* (pp. 184–201). New York, NY: Routledge.

Bråten, I., Brante, E., & Strømsø, H. (2019). Teaching sourcing in upper-secondary school: A comprehensive intervention with follow-up data. *Reading Research Quarterly, 54*(4), 481–505. https://doi.org/10.1002/rrq.253.

Bråten, I., McCrudden, M. T., Lund, E. S., Brante, E. W., & Strømsø, H. I. (2018a). 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. (2018b). The role of sourcing in discourse comprehension. In M. F. Schober, D. N. Rapp, & M. A. Britt (Eds.), *The Routledge handbook of discourse processes* (2nd ed., pp. 141–166). New York, NY: Routledge.

Britt, M. A., & Aglinskas, 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.

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.

Brozo, W. G., Sulkunen, S., Shiel, G., Garbe, C., Pandian, A., & Valtin, R. (2014). Reading, gender, and engagement: Lessons from five PISA countries. *Journal Of Adolescent & Adult Literacy, 57*(7), 584–593. https://doi.org/10.1002/jaal.291.

Choi, H. J., & Johnson, S. D. (2005). The effect of context-based video instruction on learning and motivation in online courses. *American Journal of Distance Education, 19*(4), 215–227. https://doi.org/10.1207/s15389286ajde1904_3.

Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement, 20*(1), 37–46. https://doi.org/10.1177/001316446002000104.

Coiro, J. (2011). Talking about reading as thinking: Modeling the hidden complexities of online reading comprehension. *Theory Into Practice, 50*(2), 107–115. https://doi.org/10.1080/00405841.2011.558435.

Coiro, J., Coscarelli, C., Maykel, C., & Forzani, E. (2015). Investigating criteria that seventh graders use to evaluate the quality of online information. *Journal Of Adolescent & Adult Literacy, 59*(3), 287–297. https://doi.org/10.1002/jaal.448.

EU Council recommendation on key competences for lifelong learning. (2018). The Council of the European Union. https://ec.europa.eu/education/education-in-the-eu/council-recommendation-on-key-competences-for-lifelong-learning_en. (Accessed 12 August 2019).

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.

Davey, B. (1983). Think aloud: Modeling the cognitive processes of reading comprehension. *Journal of Reading, 27*(1), 44–47.

Enders, C. K. (2010). *Applied missing data analysis*. New York: Guilford Publications.

Finnish National Board of Education. (2016). New national core curriculum for basic education: Focus on school culture and integrative approach. https://www.oph.fi/sites/default/files/documents/new-national-core-curriculum-for- basic-education.pdf. (Accessed 12 February 2020).

Fisher, D., & Frey, N. (2015). Teaching modeling using complex informational texts. *The Reading Teacher, 69*(1), 63–69. https://doi.org/10.1002/trtr.1372.

Forzani, E. (2018). How well can students evaluate online science information? Contributions of prior knowledge, gender, socioeconomic status, and offline reading ability. *Reading Research Quarterly, 53*(4), 385–390. https://doi.org/10.1002/rrq.218.

Forzani, E. (2019). A three-tiered framework for proactive critical evaluation during online inquiry. *Journal of Adolescent & Adult Literacy, 63*(4), 401–414. https://doi.org/10.1002/jaal.1004.

Frerejean, J., van Strien, J. L. H., Kirschner, P. A., & Brand-Gruwel, S. (2018). Effects of a modelling example for teaching information problem solving skills. *Journal Of Computer Assisted Learning, 34*(6), 688–700. https://doi.org/10.1111/jcal.12276.

Gardner, W., Mulvey, E. P., & Shaw, E. C. (1995). Regression analyses of counts and rates: Poisson, overdispersed Poisson, and negative binomial models. *Psychological Bulletin, 118*(3), 392–404. https://doi.org/10.1037/0033-2909.118.3.392.

van Gog, T., Verveer, I., & Verveer, L. (2014). Learning from video modeling examples: Effects of seeing the human model's face. *Computers & Education, 72*, 323–327. https://doi.org/10.1016/j.compedu.2013.12.004.

Goldman, S. R., Britt, M. A., Brown, W., Cribb, G., George, M., Greenleaf, C., et al. (2016). Disciplinary literacies and learning to read for understanding: A conceptual framework for disciplinary literacy. *Educational Psychologist, 52*(2), 219–246. https://doi.org/10.1080/00461520.2016.1168741.

González-Ibáñez, R., Gacitúa, D., Sormunen, E., & Kiili, C. (2017). Neurone: oNlinE inqUiRy experimentatiON systEm. *Proc. Assoc. Info. Sci. Technol., 54*(1), 687–689. https://doi.org/10.1002/pra2.2017.14505401117.

Harju-Luukkainen, H., Vettenranta, J., Ouakrim-Soivio, N., & Bernelius, V. (2016). Differences between students' PISA reading literacy scores and grading for mother tongue and literature at school: A geostatistical analysis of the Finnish PISA 2009 data. *Education Inquiry, 7*(4), 463–479. https://doi.org/10.3402/edui.v7.29413.

Heck, R. H. (2001). Multilevel modeling with SEM. In G. A. Marcoulides, & R. E. Schumacker (Eds.), *New developments and techniques in structural equation modelling* (pp. 89–127). Mahwah, NJ: Lawrence Erlbaum.

Hoeken, H. (2001). Anecdotal, statistical, and causal evidence: Their perceived and actual persuasiveness. *Argumentation, 15*(4), 425–437. https://doi.org/10.1023/A:1012075630523.

Holopainen, L., Kairaluoma, L., Nevala, J., Ahonen, T., & Aro, M. (2004). *Lukivaikeuksien seulontatesti nuorille ja aikuisille [Dyslexia screening test for youth and adults]*. Jyväskylä, Finland: Niilo Mäki Institute.

Judd, V. C., Farrow, L. I., & Tims, B. J. (2006). Evaluating public web site information: A process and an instrument. *Reference Services Review, 34*(1), 12–32. https://doi.org/10.1108/00907320510631571.

Kajamies, A. (2017). *Towards optimal scaffolding of low achievers' learning: Combining intertwined, dynamic, and multi-domain perspectives*. Finland: Painosalama Oy: Doctoral dissertation, University of Turku.

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.

Kanniainen, L., Kiili, C., Tolvanen, A., Aro, M., & Leppänen, P. H. T. (2019). Literacy skills and online research and comprehension: Struggling readers face difficulties online. *Reading and Writing, 32*, 2201–2222. https://doi.org/10.1007/s11145-019-09944-9.

Kiili, C., Brante, E. W., Räikkönen, E., & Coiro, J. (2020). Citing as a sourcing practice: Students' citing self-selected online sources in their essays. *Journal for the Study of Education and Development*. Advanced online publication https://www.tandfonline.com/doi/full/10.1080/02103702.2019.1690839.

Kiili, C., Bråten, I., Kullberg, N., & Leppänen, P. H. (2020a). Investigating elementary school students' text-based argumentation with multiple online information resources. *Computers & Education, 147*. https://doi.org/10.1016/j.compedu.2019.103785.

Kiili, C., Coiro, J., & Räikkönen, E. (2019). Students' evaluation of online texts during online inquiry: Working individually or in pairs. *Australian Journal of Language and Literacy, 42*, 167–183.

Kiili, C., Leu, D. J., Marttunen, M., Hautala, J., & Leppänen, P. H. T. (2018). Exploring early adolescents' evaluation of academic and commercial online resources related to health. *Reading and Writing, 31*(3), 533–557. https://doi.org/10.1007/s11145-017- 9797-2.

Kingsley, T. L., Cassady, J. C., & Tancock, S. M. (2015). Successfully promoting 21st century online research skills: Interventions in 5th-grade classrooms. *Reading Horizons, 54*(2), 91–135. https://scholarworks.wmich.edu/reading_horizons/vol54/iss2/5.

Kroustallaki, D., Kokkinaki, T., Sideridis, G. D., & Simos, P. G. (2015). Exploring students' affect and achievement goals in the context of an intervention to improve web searching skills. *Computers in Human Behavior, 49*, 156–170. https://doi.org/10.1016/j.chb.2015.02.060.

Larson, A. A., Britt, M. A., & Kurby, C. A. (2009). Improving students' evaluation of informal arguments. *The Journal of Experimental Education, 77*(4), 339–366. https://doi.org/10.3200/JEXE.77.4.339-366.

Leu, D. J., Forzani, E., Rhoads, C., Maykel, C., Kennedy, C., & Timbrell, N. (2015). The new literacies of online research and comprehension: Rethinking the reading achievement gap. *Reading Research Quarterly, 50*(1), 37–59. https://doi.org/10.1002/rrq.85.

Livingstone, S., Mascheroni, G., & Staksrud, E. (2018). European research on children's Internet use: Assessing the past and anticipating the future. *New Media & Society, 20*(3), 1103–1122. https://doi.org/10.1177/1461444816685930.

Macedo-Rouet, M., Braasch, J. L. G., Britt, M. A., & Rouet, J.-F. (2013). Teaching fourth and fifth graders to evaluate information sources during text comprehension. *Cognition and Instruction, 31*(2), 204–226. https://doi.org/10.1080/07370008.2013.769995.

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

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.

Mateos, M., Martín, E., Cuevas, I., Villalón, R., Martínez, I., & González-Lamas, J. (2018). Improving written argumentative synthesis by teaching the integration of conflicting information from multiple sources. *Cognition and Instruction, 36*(2), 119–138. https://doi.org/10.1080/07370008.2018.1425300.

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.

Means, M. L., & Voss, J. F. (1996). Who reasons well? Two studies of informal reasoning among children of different grade, ability, and knowledge levels. *Cognition and Instruction, 14*(2), 139–178. https://doi.org/10.1207/s1532690xci1402_1.

Metzger, M. (2007). Making sense of credibility on the Web: Models for evaluating online information and recommendations for future research. *Journal Of the American Society For Information Science and Technology, 58*(13), 2078–2091. https://doi.org/10.1002/asi.20672.

Morris, S. B., & DeShon, R. P. (2002). Combining effect size estimates in meta-analysis with repeated measures and independent-groups designs. *Psychological Methods, 7*(1), 105–125. https://doi.org/10.1037//1082-989X.7.1.105.

Muthén, B. O. (1991). Multilevel factor analysis of class and student achievement components. *Journal of Educational Measurement, 28*(4), 338–354. https://doi.org/10.1111/j.1745-3984.1991.tb00363.x.

Muthén, B. O. (1997). Latent variable modeling of longitudinal and multilevel data. *Sociological Methodology, 27*, 453–480.

Muthén, L., & Muthén, B. O. (1998–2017). *Mplus user's guide* (8th ed). Los Angeles, CA: Muthén & Muthén.

Paul, J., Macedo-Rouet, M., Rouet, J.-F., & Stadtler, M. (2017). Why attend to source information when reading online? The perspective of ninth grade students from two different countries. *Computers & Education, 113*, 339–354. https://doi.org/10.1016/j.compedu.2017.05.020.

Pérez, A., Potocki, A., Stadtler, M., Macedo-Rouet, M., Paul, J., Salmerón, L., et al. (2018). Fostering teenagers' assessment of information reliability: Effects of a classroom intervention focused on critical source dimensions. *Learning and Instruction, 58*, 53–64. 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). Mahwah, NJ: Erlbaum.

Pressley, M., & Afflerbach, P. (1995). *Verbal protocols of reading: The nature of constructively responsive reading*. New York, NY: Routledge.

Quintana, C., Zhang, M., & Krajcik, J. S. (2005). A framework for supporting metacognitive aspects of online inquiry through software-based scaffolding. *Educational Psychologist, 40*(4), 235–244. https://doi.org/10.1207/s15326985ep4004_5.

Shanahan, T., & Shanahan, C. (2012). What is disciplinary literacy and why does it matter? *Topics in Language Disorders, 32*(1), 7–18. https://doi.org/10.1097/TLD.0b013e318244557a.

Sormunen, E., González-Ibáñez, R., Kiili, C., Leppänen, P. H. T., Mikkilä-Erdmann, M., Erdmann, N., et al. (2018). A performance-based test for assessing students' online inquiry competences in schools. In S. Kurbanoğlu, J. Boustany, S. Špiranec, E. Grassian, D. Mizrachi, & L. Roy (Eds.), *Information literacy in the workplace. ECIL, 2017. Communications in computer and information science, 810* pp. 673–682).

Stadtler, M., & Bromme, R. (2014). The content–source integration model: A taxonomic description of how readers comprehend conflicting scientific information. In D. Rapp, & J. Braasch (Eds.), *Processing inaccurate information: Theoretical and applied perspectives from cognitive science and the educational sciences* (pp. 379–402). MIT Press.

Teasley, S. D. (1995). The role of talk in children's peer collaborations. *Developmental Psychology, 31*(2), 207–220. https://doi.org/10.1037/0012-1649.31.2.207.

Walraven, A., Brand-Gruwel, S., & Boshuizen, H. P. A. (2010). Fostering transfer of websearchers' evaluation skills: A field test of two transfer theories. *Computers in Human Behavior, 26*(4), 716–728. https://doi.org/10.1016/j.chb.2010.01.008.

Walraven, A., Brand-Gruwel, S., & Boshuizen, H. P. A. (2013). Fostering students' evaluation behaviour while searching the Internet. *Instructional Science, 41*(1), 125–146. https://doi.org/10.1007/s11251-012-9221-x.

van Wermeskerken, M., & van Gog, T. (2017). Seeing the instructor's face and gaze in demonstration video examples affects attention allocation but not learning. *Computers & Education, 113*, 98–107. https://doi.org/10.1016/j.compedu.2017.05.013.

Zhang, S., & Duke, N. K. (2011). The impact of instruction in the WWWDOT framework on student's disposition and ability to evaluate web sites as sources of information. *The Elementary School Journal, 112*(1), 132–154. https://doi.org/10.1086/660687.

Zhang, W.-X., Hsu, Y.-S., Wang, C.-Y., & Ho, Y.-T. (2015). Exploring the impacts of cognitive and metacognitive prompting on students' scientific inquiry practices within an e- learning environment. *International Journal of Science Education, 37*(3), 529–553. https://doi.org/10.1080/09500693.2014.996796.

II

STUDENTS' ABILITIES TO EVALUATE THE CREDIBILITY OF ONLINE TEXTS: THE ROLE OF INTERNET-SPECIFIC EPISTEMIC JUSTIFICATIONS

by

Elina K. Hämäläinen, Carita Kiili, Eija Räikkönen, & Miika Marttunen, 2021

Journal of Computer Assisted Learning, 37(5), 1409-1422.

https://doi.org/10.1111/jcal.12580

Received: 26 June 2020 | Revised: 27 April 2021 | Accepted: 23 May 2021

DOI: 10.1111/jcal.12580

ARTICLE

Journal of Computer Assisted Learning WILEY

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

Elina K. Hämäläinen[1] | Carita Kiili[2,3] | Eija Räikkönen[4] | Miika Marttunen[1]

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

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

[3]Department of Education, University of Oslo, Norway

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

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

Abstract

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

KEYWORDS

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

1 | INTRODUCTION

The current COVID-19 pandemic has challenged publics' abilities to evaluate the credibility of health information online. Misleading information has spread rapidly via the Internet. Moreover, experts may disagree in a novel uncertain situation where it takes time for scientific research to yield results. As a whole, the current online debate reflects a post-truth world in which laypersons may disagree about evidence-based facts and place more weight on their personal beliefs than on scientific knowledge when deciding what to believe (Sinatra & Lombardi, 2020).

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

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

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

1.1 | Theoretical frameworks

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

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

The theory of multiple documents comprehension (Britt et al., 2018; Perfetti et al., 1999; Rouet, 2006) describes how, to achieve their reading goals, readers select, evaluate and use information from more than one document. Compared to single document comprehension, during which readers integrate text contents with their prior knowledge, multiple document comprehension presents additional challenges in building a coherent representation of the information contained in different documents. For example, contradictory information gathered from different sources might be difficult to integrate coherently. To address these challenges, the documents model framework proposes that readers need to form two representations: an integrated mental model and an intertext model. The integrated mental model refers to the representation of contents across the documents organized in accordance with the reading task. The intertext model, in turn, refers to the representation of source information (e.g., authors' credentials and intentions) and links between the sources to its content and rhetorical relationships between the sources. By combining these models, readers can understand complex and potentially conflicting information by incorporating the contents of documents into their respective sources.

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

1.2 | Evaluation of credibility

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

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

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

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

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

1.3 | Justifications for knowing

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

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

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

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

1.4 | The present study

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

The specific research questions were:

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

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

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

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

2 | METHOD

2.1 | Participants

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

2.2 | Online inquiry task

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

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

After reading the task scenario, the online inquiry task proceeded in four phases (Leu et al., 2019): (1) considering information need to solve the problem; (2) locating information with a search engine to select three online texts; (3) identifying main ideas of each selected text and evaluating the credibility of the texts and (4) writing a justified recommendation. Each task phase began on a separate page. Students were able to move between the task phases by using forward and backward buttons. The data of this study originates from Task Phases 2 (selections) and 3 (credibility evaluations).

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

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

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

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

2.3 | Other measures

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

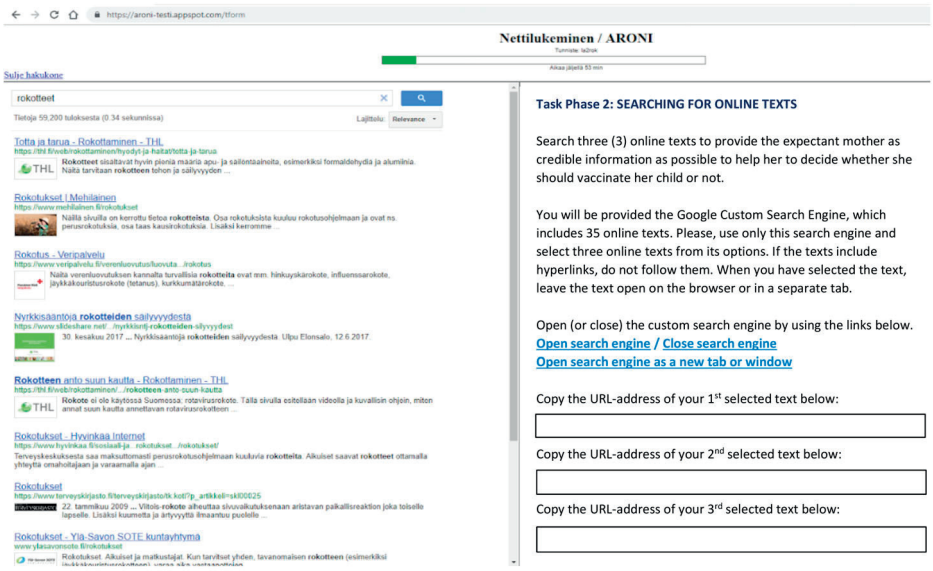


FIGURE 1 Task phase 2: Locating and selecting online texts [Colour figure can be viewed at wileyonlinelibrary.com]

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

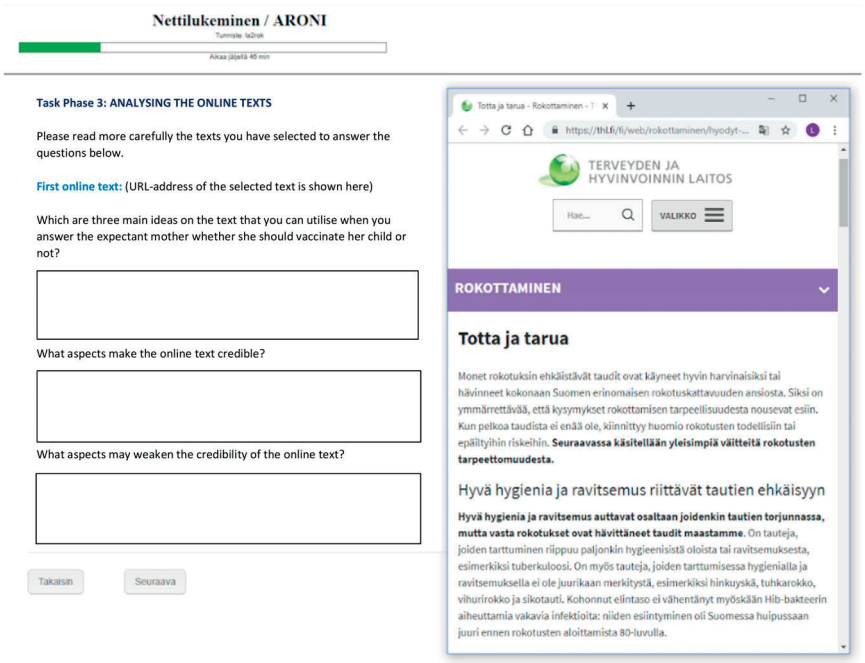


FIGURE 2 Task phase 3: Identifying Main ideas and evaluating the credibility of online texts [Colour figure can be viewed at wileyonlinelibrary.com]

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

I evaluate it in relation to my knowledge of this topic'), Justification by Authority (e.g., 'When I read information from the Internet related to my school task, I evaluate whether this information is written by an expert') and Justification by Multiple Sources (e.g., 'To determine whether the information related to my school task I find on the Internet is trustworthy, I compare information from multiple sources'). Instead of using the original 10-point scale, we used a 5-point scale with labels: 1 = strongly disagree, 2 = partly disagree, 3 = not disagree or agree, 4 = partly agree, 5 = strongly agree. Thus, the ISEJ-items were measured on the ordinal level and used as approximations of students' continuous level beliefs in justifications for knowing.

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

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

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

2.4 | Procedure

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

2.5 | Data analysis

2.5.1 | Selection of online texts

In Task Phase 2 (Figure 1), students selected three online texts by using Google Custom Search Engine that included 35 pre-selected texts per topic. Although instructed to select only texts included in the custom search engine, one-fourth of students also selected other texts, mostly one. These other texts ($N = 64$) accounted for 11% of all selected texts (Table 1). Almost 60% of these texts appeared in the same venue as the pre-selected texts suggesting that students probably navigated within the website. We incorporated the other selected texts into the original textbase and used the same rubric to score all 134 texts.

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

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

2.5.2 | Students' credibility evaluations

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

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

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

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

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

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

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

Step 2: Evaluation performance. To assess students' evaluation performance across three online texts, we created a scoring rubric that utilized the analysis conducted in Step 1. The scoring rubric, presented in Table 3, acknowledged different credibility aspects and depth in students' reasoning (justifications at the highest, 3 points level). The scoring rubric reflected whether students' responses across the three texts demonstrated their abilities to evaluate different credibility aspects and engage in deep reasoning (See also Kiili et al., 2019). In other words, students had three possibilities to evaluate each aspect, and they were given credit in the scoring system if they evaluated the aspect at least once. By this procedure, we tried to minimize the effect of the evaluation of different text combinations.

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

2.5.3 | Statistical analyses

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

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

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

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

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

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

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

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

Cholesky factoring for the justification for knowing dimensions (Figure 3) was performed so that we set the first Cholesky factor (labelled 'PJ: Cholesky') to explain all the variance unique to the Personal Justification dimension and the variance it shares with the other two dimensions. Then, we set the second Cholesky factor (labelled 'JA: Cholesky') to explain the unique variance of the Justification by Authority dimension and the variance it shares with Justification by Multiple Sources. The third Cholesky factor (labelled 'JMS: Cholesky') captured the remaining (unique) variance of the JMS dimension. The correlations between the Cholesky factors and the correlations between the original justification for knowing dimensions and their cross-correlations were fixed to 0.

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

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

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

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

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

3 | RESULTS

3.1 | Credibility evaluations

3.1.1 | Evaluation of credibility aspects

Table 4 shows that students most often evaluated the venue and evidence presented in online texts. Specifically, almost 90% of students evaluated the venue and over 75% the evidence at least once across three online texts. In contrast, students only sparsely evaluated intentions or applied corroboration as a credibility evaluation criterion. Furthermore, the students most often reached the highest level in their evaluations when they evaluated the evidence or venue. Over one-fourth of the students evaluated the evidence at least once at the highest level across three texts and correspondingly, one-fifth of the students when evaluating the venue.

3.1.2 | Evaluation performance

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

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

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

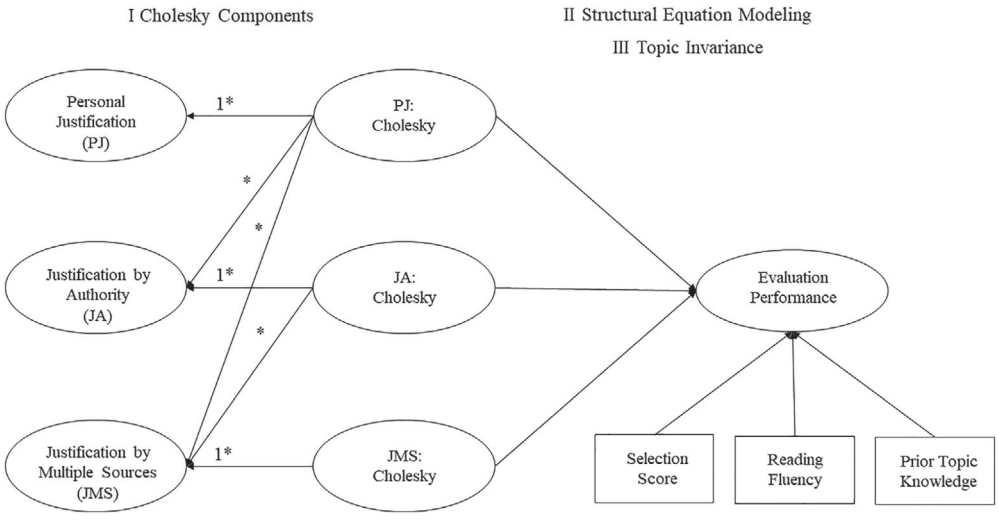


FIGURE 3 Conceptual model of the relationships between ISEJ-factors, evaluation performance and control variables. 1* fixed to one, * freely estimated

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

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

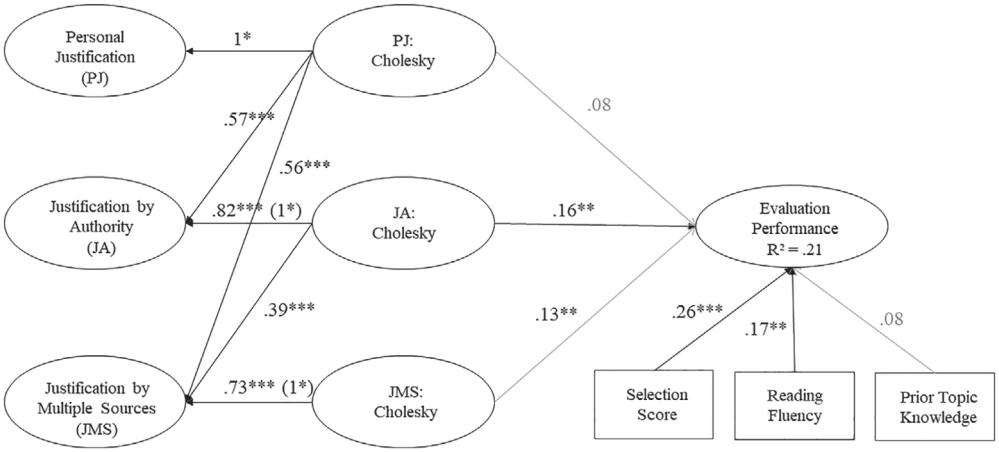


FIGURE 4 Relations between ISEJ-factors, evaluation performance and control variables. Statistically significant standardized estimates (**p < 0.01, ***p < 0.001) are written in black (nonsignificant estimates written in grey). 1* fixed to one

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

4 | DISCUSSION

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

The results showed considerable variation in students' abilities to evaluate the credibility of online texts, a finding also previously reported (e.g., Kiili et al., 2019; McGrew et al., 2018). Over one-third of the students demonstrated a high ability to move across different credibility aspects with some deep-level justifications when evaluating the credibility across three online texts. The ability to pay attention to different aspects of credibility provides students with flexibility in their evaluations. Further, paying attention to multiple aspects of credibility is important, as an accurate evaluation often cannot be made by relying on one aspect alone (Forzani, 2020). Almost one-tenth of students performed very poorly, relying at most on only one aspect of credibility. Additionally, 20% of students demonstrated limited abilities to engage in versatile and sophisticated evaluation. This is worrying, as adolescents with poor evaluation skills may be particularly vulnerable to mis- and disinformation.

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

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

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

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

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

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

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

4.1 | Limitations and future research

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

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

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

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

4.2 | Instructional implications

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

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

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

ACKNOWLEDGMENTS

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

CONFLICT OF INTEREST

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

PEER REVIEW

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

DATA AVAILABILITY STATEMENT

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

ORCID

Elina K. Hämäläinen https://orcid.org/0000-0001-7561-0530
Carita Kiili https://orcid.org/0000-0001-9189-4094
Eija Räikkönen https://orcid.org/0000-0003-4450-9178
Miika Marttunen https://orcid.org/0000-0002-4554-9764

REFERENCES

Braasch, J. L. G., Bråten, I., Strømsø, H. I., Anmarkrud, Ø., & Ferguson, L. E. (2013). Promoting secondary school students' evaluation of source features of multiple documents. *Contemporary Educational Psychology*, *38*, 180–195. https://doi.org/10.1016/j.cedpsych.2013.03.003

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

Bråten, I., Brandmo, C., & Kammerer, Y. (2019). A validation study of the Internet-Specific Epistemic Justification Inventory with Norwegian preservice teachers. *Journal of Educational Computing Research*, *57*, 877–900. https://doi.org/10.1177/0735633118769438

Bråten, I., Britt, M. A., Strømsø, H. I., & Rouet, J.-F. (2011). The role of epistemic beliefs in the comprehension of multiple expository texts: Toward an integrated model. *Educational Psychologist*, *46*, 48–70. https://doi.org/10.1080/00461520.2011.538647

Bråten, I., McCrudden, M. T., Stang Lund, E., Brante, E. W., & Strømsø, H. I. (2018). Task-oriented learning with multiple documents. Effects of topic familiarity, author expertise, and content relevance on document selection, processing, and use. *Reading Research Quarterly*, *53*, 345–365. https://doi.org/10.1002/rrq.197

Bråten, I., Stadtler, M., & Salmerón, L. (2018). The role of sourcing in discourse comprehension. In M. F. Schober, D. N. Rapp, & M. A. Britt (Eds.), *Handbook of discourse processes* (2nd ed.). Routledge.

Bråten, I., Strømsø, H. I., & Samuelstuen, M. S. (2005). The relationship between internet-specific epistemological beliefs and learning within internet technologies. *Journal of Educational Computing Research*, *33*, 141–171. https://doi.org/10.2190/E763-X0LN-6NMF-CB86

Britt, M. A., Richter, T., & Rouet, J. F. (2014). Scientific literacy: The role of goal-directed reading and evaluation in understanding scientific information. *Educational Psychologist*, *4*, 104–122. https://doi.org/10.1080/00461520.2014.916217

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

Bromme, R., & Goldman, S. R. (2014). The public's bounded understanding of science. *Educational Psychologist*, *49*, 59–69. https://doi.org/10.1080/00461520.2014.921572

Cartiff, B. M., Duke, R. F., & Greene, J. A. (2020). The effect of epistemic cognition interventions on academic achievement: A meta-analysis. *Journal of Educational Psychology*, *113*, 477–498. https://doi.org/10.1037/edu0000490

Chen, F. F. (2007). Sensitivity of goodness of fit indexes to lack of measurement invariance. *Structural Equation Modeling*, *14*, 464–504. https://doi.org/10.1080/10705510701301834

Coiro, J., Coscarelli, C., Maykel, C., & Forzani, E. (2015). Investigating criteria that seventh graders use to evaluate the quality of online information. *Journal of Adolescent & Adult Literacy*, *59*, 287–297. https://doi.org/10.1002/jaal.448

de Jong, P. F., & van der Leij, A. (1999). Specific contributions of phonological abilities to early reading acquisition: Results from a Dutch latent variable longitudinal study. *Journal of Educational Psychology*, *91*, 450–476. https://doi.org/10.1037/0022-0663.91.3.450

Enders, C. K. (2010). *Applied missing data analysis*. Guilford Publications.

Ferguson, L. E., Bråten, I., & Strømsø, H. I. (2012). Epistemic cognition when students read multiple documents containing conflicting scientific evidence: A think-aloud study. *Learning and Instruction*, *22*, 103–120. https://doi.org/10.1016/j.learninstruc.2011.08.002

Forzani, E. (2018). How well can students evaluate online science information? Contributions of prior knowledge, gender, socioeconomic status, and offline reading ability. *Reading Research Quarterly*, *53*, 385–390. https://doi.org/10.1002/rrq.218

Forzani, E. (2020). A three-tiered framework for proactive critical evaluation during online inquiry. *Journal of Adolescent & Adult Literacy*, *63*, 401–414. https://doi.org/10.1002/jaal.1004

Freeman, J. L., Caldwell, P. H., & Scott, K. M. (2020). The role of trust when adolescents search for and appraise online health information. *The Journal of Pediatrics*, *221*, 215–223. https://doi.org/10.1016/j.jpeds.2020.02.074

Gerjets, P., Kammerer, Y., & Werner, B. (2011). Measuring spontaneous and instructed evaluation processes during web search: Integrating concurrent thinking-aloud protocols and eye-tracking data. *Learning and Instruction*, *21*, 220–231. https://doi.org/10.1016/j.learninstruc.2010.02.005

Greene, J. A., Azevedo, R., & Torney-Purta, J. (2008). Modeling epistemic and ontological cognition: Philosophical perspectives and

methodological directions. *Educational Psychologist*, *43*, 142–160. https://doi.org/10.1080/00461520802178458

Greene, J. A., Cartiff, B. M., Duke, R. F., & Deekens, V. M. (2019). A nation of curators. Educating students to be critical consumers and users of online information. In P. Kendeou, D. H. Robinson, & M. T. McCrudden (Eds.), *Misinformation and fake news in education* (pp. 187–206). Information Age Publishing.

Greene, J. A., & Yu, S. B. (2015). Educating critical thinkers: The role of epistemic cognition. *Policy Insights From the Behavioral and Brain Sciences*, *3*, 45–53. https://doi.org/10.1177/2372732215622223

Hämäläinen, E. K., Kiili, C., Marttunen, M., Räikkönen, E., González-Ibáñez, R., & Leppänen, P. H. (2020). Promoting sixth graders' credibility evaluation of web pages: An intervention study. *Computers in Human Behaviour*, *110*, 106372. https://doi.org/10.1016/j.chb.2020.106372

Hautala, J., Kiili, C., Kammerer, Y., Loberg, O., Hokkanen, S., & Leppänen, P. H. (2018). Sixth graders' evaluation strategies when reading internet search results: An eye-tracking study. *Behaviour & Information Technology*, *37*, 761–773. https://doi.org/10.1080/0144929X.2018.1477992

Hoeken, H. (2001). Anecdotal, statistical, and causal evidence: Their perceived and actual persuasiveness. *Argumentation*, *15*, 425–437. https://doi.org/10.1023/A:1012075630523

Hofer, B. K., & Pintrich, P. R. (1997). The development of epistemological theories: Beliefs about knowledge and knowing and their relation to learning. *Review of Educational Research*, *67*, 88–140. https://doi.org/10.3102/00346543067001088

Holopainen, L., Kairaluoma, L., Nevala, J., Ahonen, T., & Aro, M. (2004). *Lukivaikeuksien seulontatesti nuorille ja aikuisille [Dyslexia screening test for youth and adults]*. Niilo Mäki Institute.

Hu, L. T., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling*, *6*, 1–55. https://doi.org/10.1080/10705519909540118

Jeong, H., & Hmelo-Silver, C. E. (2016). Seven affordances of computer-supported collaborative learning: How to support collaborative learning? How can technologies help? *Educational Psychologist*, *51*, 247–265. https://doi.org/10.1080/00461520.2016.1158654

Kammerer, Y., Amann, D. G., & Gerjets, P. (2015). When adults without university education search the internet for health information: The roles of internet-specific epistemic beliefs and a source evaluation intervention. *Computers in Human Behavior*, *48*, 297–309. https://doi.org/10.1016/j.chb.2015.01.045

Kammerer, Y., Bråten, I., Gerjets, P., & Strømsø, H. I. (2013). The role of internet-specific epistemic beliefs in laypersons' source evaluations and decisions during web search on a medical issue. *Computers in Human Behavior*, *29*, 1193–1203. https://doi.org/10.1016/j.chb.2012.10.012

Kammerer, Y., Gottschling, S., & Bråten, I. (2021). The role of internet-specific justification beliefs in source evaluation and corroboration during web search on an unsettled socio-scientific issue. *Journal of Educational Computing Research*, *59*, 42–378. https://doi.org/10.1177/0735633120952731

Kammerer, Y., Kalbfell, E., & Gerjets, P. (2016). Is this information source commercially biased? How contradictions between web pages stimulate the consideration of source information. *Discourse Processes*, *53*, 430–456. https://doi.org/10.1080/0163853X.2016.1169968

Kanniainen, L., Kiili, C., Tolvanen, A., Aro, M., & Leppänen, P. H. T. (2019). Literacy skills and online research and comprehension: Struggling readers face difficulties online. *Reading and Writing*, *32*, 2201–2222. https://doi.org/10.1007/s11145-019-09944-9

Kiili, C., Coiro, J., & Räikkönen, E. (2019). Students' evaluation of information during online inquiry: Working individually or in pairs. *Australian Journal of Language and Literacy*, *42*, 167–183.

Kiili, C., Leu, D. J., Marttunen, M., Hautala, J., & Leppänen, P. H. T. (2018). Exploring early adolescents evaluation of academic and commercial online resources related to health. *Reading and Writing*, *31*, 533–557. https://doi.org/10.1007/s11145-017-9797-2

Knight, S., Rienties, B., Littleton, K., Mitsui, M., & Tempelaar, D. (2017). The relationship of (perceived) epistemic cognition to interaction with resources on the internet. *Computers in Human Behavior*, *73*, 507–518. https://doi.org/10.1016/j.chb.2017.04.014

Kohnen, A. M., & Mertens, G. E. (2019). "I'm always kind of double-checking": Exploring the information-seeking identities of expert generalists. *Reading Research Quarterly*, *54*, 279–297. https://doi.org/10.1002/rrq.245

Leu, D. J., Kinzer, C. K., Coiro, J., Castek, J., & Henry, L. A. (2019). New literacies: A dual level theory of the changing nature of literacy, instruction, and assessment. In D. E. Alvermann, N. J. Unrau, M. Sailors, & R. B. Ruddell (Eds.), *Theoretical models and processes of literacy* (7th ed., pp. 319–346). Taylor & Francis.

Little, R. J. (1988). A test of missing completely at random for multivariate data with missing values. *Journal of the American Statistical Association*, *83*, 1198–1202. https://doi.org/10.1080/01621459.1988.10478722

McCrudden, M. T. (2018). Text relevance and multiple-source use. In J. L. G. Braasch, I. Bråten, & M. T. McCrudden (Eds.), *Handbook of multiple source use* (pp. 168–183). Routledge.

McGrew, S., Breakstone, J., Ortega, T., Smith, M., & Wineburg, S. (2018). Can students evaluate online sources? Learning from assessments of civic online reasoning. *Theory & Research in Social Education*, *46*, 165–193. https://doi.org/10.1080/00933104.2017.1416320

Meola, M. (2004). Chucking the checklist: A contextual approach to teaching undergraduates web-site evaluation. *Libraries and the Academy*, *4*, 331–344. https://doi.org/10.1353/pla.2004.0055

Meredith, W. (1993). Measurement invariance, factor analysis and factorial invariance. *Psychometrika*, *58*, 525–543. https://doi.org/10.1007/BF02294825

Muthén, L., & Muthén, B. O. (1998–2017). *Mplus user's guide* (8th ed.). Muthén & Muthén.

Pan, B., Hembrooke, H., Joachims, T., Lorigo, L., Gay, G., & Granka, L. (2007). In Google we trust: Users' decisions on rank, position, and relevance. *Journal of Computer-Mediated Communication*, *12*, 801–823. https://doi.org/10.1111/j.1083-6101.2007.00351.x

Paul, J., Macedo-Rouet, M., Rouet, J. F., & Stadtler, M. (2017). Why attend to source information when reading online? The perspective of ninth grade students from two different countries. *Computers & Education*, *113*, 339–354. https://doi.org/10.1016/j.compedu.2017.05.020

Perfetti, C. A., Rouet, J.-F., & Britt, M. A. (1999). Towards a theory of documents representation. In H. van Oostendorp & S. Goldman (Eds.), *The construction of mental representations during reading* (pp. 99–122). Erlbaum.

Potocki, A., de Pereyra, G., Ros, C., Macedo-Rouet, M., Stadtler, M., Salmerón, L., & Rouet, J.-F. (2020). The development of source evaluation skills during adolescence: Exploring different levels of source processing and their relationships. *Journal for the Study of Education and Development*, *43*, 19–59. https://doi.org/10.1080/02103702.2019.1690848

Raykov, T., Dimitrov, D. M., & Asparouhov, T. (2010). Evaluation of scale reliability with binary measures using latent variable modeling. *Structural Equation Modeling*, *17*, 265–279. https://doi.org/10.1080/10705511003659417

Rieh, S. Y. (2002). Judgment of information quality and cognitive authority in the web. *Journal of the American Society for Information Science and Technology*, *53*, 145–161. https://doi.org/10.1002/asi.10017

Rouet, J. F. (2006). *The skills of document use: From text comprehension to web-based learning*. Erlbaum.

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

Salmerón, L., Strømsø, H. I., Kammerer, Y., Stadtler, M., & van den Broek, P. (2018). Comprehension processes in digital reading. In M. Barzillai, J. Thomson, S. Schroeder, & P. van den Broek (Eds.), *Learning*

to read in a digital world (pp. 91–120). Studies in written language and literacy 17). John Benjamins Publishing Company.

Sandoval, W. A., Sodian, B., Koerber, S., & Wong, J. (2014). Developing children's early competencies to engage with science. *Educational Psychologist*, *49*, 39–152. https://doi.org/10.1080/00461520.2014.917589

Satorra, A., & Bentler, P. M. (2001). A scaled difference chi-square test statistic for moment structure analysis. *Psychometrika*, *66*, 507–514. https://doi.org/10.1007/BF02296192

Scharrer, L., Stadtler, M., & Bromme, R. (2019). Judging scientific information: Does source evaluation prevent the seductive effect of text easiness? *Learning and Instruction*, *63*, 101215. https://doi.org/10.1016/j.learninstruc.2019.101215

Sinatra, G. M., & Lombardi, D. (2020). Evaluating sources of scientific evidence and claims in the post-truth era may require reappraising plausibility judgments. *Educational Psychologist*, *55*, 120–131. https://doi.org/10.1080/00461520.2020.1730181

Strømsø, H. I., & Bråten, I. (2010). The role of personal epistemology in the self-regulation of internet-based learning. *Metacognition Learning*, *5*, 91–111. https://doi.org/10.1007/s11409-009-9043-7

Wineburg, S. (1991). Historical problem solving: A study of the cognitive processes used in the evaluation of documentary and pictorial evidence. *Journal of Educational Psychology*, *83*, 73–87. https://doi.org/10.1037/0022-0663.83.1.73

SUPPORTING INFORMATION

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

How to cite this article: Hämäläinen, E. K., Kiili, C., 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*, 1–14. https://doi.org/10.1111/jcal.12580

III

TEACHING SOURCING DURING ONLINE INQUIRY – ADOLESCENTS WITH THE WEAKEST SKILLS BENEFITED THE MOST

by

Elina K. Hämäläinen, Carita Kiili, Eija Räikkönen, Minna Lakkala, Liisa Ilomäki, Auli Toom, & Miika Marttunen 2023

Instructional Science, 51(1), 135-163.

https://doi.org/10.1007/s11251-022-09597-2

Instructional Science (2023) 51:135–163
https://doi.org/10.1007/s11251-022-09597-2

ORIGINAL RESEARCH

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

Elina K. Hämäläinen[1] · Carita Kiili[2] · Eija Räikkönen[3] · Minna Lakkala[4] · Liisa Ilomäki[4] · Auli Toom[5] · Miika Marttunen[1]

Received: 14 December 2021 / Accepted: 20 July 2022 / Published online: 27 October 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 (4×75 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

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

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

[3] Faculty of Education and Psychology, University of Jyväskylä, Jyväskylä, Finland

[4] Faculty of Educational Sciences, University of Helsinki, Helsinki, Finland

[5] 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 (9×90 min). In the scripted lessons (3×90 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 (3×90 min) and a group-based oral assignment (3×90 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 (4×75 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 & Aglinskas, 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_{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 (4×75 min lessons) on online inquiry as a part of their Texts and influence course (total of 23×75 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 4×75 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. Student 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., Stadtler & 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	• What kind of information do you need to advice the expectant mother on whether she should vaccinate her child (vaccination topic)? • What kind of information do you need to advice the student on whether he should avoid saturated fats (fats topic)?	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	• 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). • Search for three online texts that help you to provide the student with credible information on whether he should avoid saturated fats in his diet (fats topic).	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	• What is your position on whether the expectant mother should vaccinate her child (vaccination topic)? • What is your position on whether the student should avoid saturated fats (fats topic)? • Write below the justifications that support your position. Indicate sources that you rely on.	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 *one or two indications* of sourcing that represent source-content link, source-source link, or evaluative statement. 4 p. = student's written product includes three *indications* of sourcing that represent source-content link, source-source link, and/or evaluative statement. 5 p. = student's written product includes *at least four indications* of sourcing that represent source-content link, source-source link, and/or evaluative statement. 6 p. = student's written product includes *at least five indications* of sourcing that represent *at least two categories*: source-content link, source-source link, and/or evaluative statement. 7 p. = student's written product includes *at least five indications* of sourcing that represent *all categories*: source-content link, source-source link, and evaluative statement.	.78

Note. 10% of students' responses were coded for inter-rater relability (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 (RCI < -1.96), those who showed no change (-1.96 ≤ RCI ≤ 1.96), those who showed a reliable positive change (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 (RCI > 1.96 + 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% *CI*s 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 χ^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 fats task students.

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

	Intervention group (N=175–191)			Control group (N=143–162)		
Pre-test measures (observed range)	*M*	*SD*	*Md*	*M*	*SD*	*Md*
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 (β) and Negative Binomial Regression analysis (IRR; 95% CI) for the Associations Between Predictors, Independent Variable (Group) and Students' Sourcing Performance in the Post-Test

	Dependent variables			
Predictors	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
	β	*β*	*β*	*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] *
Group (0=control, 1=intervention)	**0.08**	**0.19*** **	**0.18****	**2.23 [1.28; 3.88]***
Cohens'*d*for group effect [95%*CI*]	**0.16 [-0.05; 0.38]**	**0.39 [0.17; 0.61]**	**0.37 [0.15; 0.58]**	
	$R^2 = 0.20^{***}$	$R^2 = 0.29^{***}$	$R^2 = 0.24^{***}$	

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 ≤ -1.96) *f* (%)	No change (-1.96≤RCI≤1.96) *f* (%)	Reliable positive change (RCI > 1.96) *f* (%)	Clear positive change (RCI > 1.96 + 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 *M (SD)*	Reading fluency (range 0–100) *M (SD)*	Prior topic knowledge (range 0–6) *M (SD)*
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' *d*		
Sourcing in specifying information need (*N* = 170–172)	1 vs. 2	**1.52 [1.27; 1.78], *d* = -1.80**	-1.79 [-8.97; 5.25]	-0.23 [-0.72; 0.27]
	1 vs. 4	**1.83 [1.39; 2.28], *d* = -3.58**	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 (*N* = 170–171)	1 vs. 2	**1.30 [1.06; 1.57], *d* = -2.31**	4.70 [-0.88; 10.48]	0.46 [-0.00; 0.86]
	1 vs. 4	**1.37 [1.16; 1.64], *d* = -2.76**	-2.81 [-9.27; 4.07]	0.05 [-0.48; 0.56]
	2 vs. 4	0.07 [-0.04; 0.19]	**-7.51 [-13.05; -2.09], *d* = 0.47**	-0.41 [-0.83; 0.03]
Sourcing in credibility judgments (*N* = 177–179)	1 vs. 2	**1.93 [0.89; 2.88], *d* = -1.27**	-1.69 [-21.06; 17.68]	0.48 [-0.26; 1.21]
	1 vs. 3	**2.93 [1.74; 4.03], *d* = -2.40**	-4.82 [-26.33; 16.69]	0.74 [-0.05; 1.54]
	1 vs. 4	**3.33 [2.07; 4.50], *d* = -2.57**	1.47 [-20.51; 23.44]	0.13 [-0.77; 1.10]
	2 vs. 3	**1.00 [0.41; 1.61], *d* = -0.67**	-3.13 [-13.79; 7.53]	0.26 [-0.23; 0.75]
	2 vs. 4	**1.40 [0.63; 2.09], *d* = -0.93**	3.16 [-8.40; 14.72]	-0.35 [-1.01; 0.32]
	3 vs. 4	0.40 [-0.47; 1.27]	6.29 [-8.59; 21.17]	-0.61 [-1.39; 0.21]
Sourcing in written product (*N* = 178–180)	1 vs. 2	**2.22 [1.38; 2.99], *d* = -1.24**	-1.68 [-11.10; 8.06]	-0.27 [-0.66; 0.09]
	1 vs. 3	**3.34 [2.39; 4.25], *d* = -2.49**	1.63 [-9.86; 13.40]	-0.28 [-0.89; 0.33]
	1 vs. 4	**4.11 [3.18; 4.89], *d* = -3.36**	4.13 [-6.34; 15.12]	-0.40 [-1.00; 0.14]
	2 vs. 3	**1.12 [0.52; 1.76], *d* = -0.64**	3.31 [-3.97; 10.64]	-0.01 [-0.60; 0.57]
	2 vs. 4	**1.89 [1.32; 2.40], *d* = -1.08**	5.82 [-0.66; 11.99]	-0.13 [-0.69; 0.41]
	3 vs. 4	**0.77 [0.10; 1.41], *d* = -0.66**	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 & Aglinskas, 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 & Aglinskas, 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.

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-Elooz, 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., & Strømsø, 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., & Strømsø, H. I. (2019). Teaching sourcing in upper secondary school: A comprehensive sourcing intervention with follow-up data. *Reading Research Quarterly, 54*(4), 481–505. https://doi.org/10.1002/rrq.253

Bråten, I., McCrudden, M. T., Stang Lund, E., Brante, E. W., & Strømsø, 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., Strømsø, 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., & Aglinskas, 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

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). *Lukivaikeuksienseulontatesti 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 ClinicalPsychology, 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., & Räikkonen, 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., Kanniainen, 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., Cassady, 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

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., Stadtler, 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., & Strømsø, 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 forupper secondary education].* Opetushallitus.

Paul, J., Cerdán, R., Rouet, J. F., & Stadtler, 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., Stadtler, M., Macedo-Rouet, M., Paul, J., Salmerón, L., & Rouet, J. F. (2018). Fostering teenagers' assessment of information reliability: Effects of a classroom intervention focused on critical source dimensions. *Learning and Instruction, 58*, 53–64. 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., Stadtler, 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/opiskt/2020/opiskt_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

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