

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

Author(s): Kiili, Carita; Coiro, Julie; Räikkönen, Eija

Title: Students' evaluation of information during online inquiry : Working individually or in pairs

Year: 2019

Version: Accepted version (Final draft)

Copyright: © The Author & ALEA, 2019

Rights: In Copyright

Rights url: <http://rightsstatements.org/page/InC/1.0/?language=en>

Please cite the original version:

Kiili, C., Coiro, J., & Räikkönen, E. (2019). Students' evaluation of information during online inquiry : Working individually or in pairs. *Australian Journal of Language and Literacy*, 42(3), 167-183. <https://doi.org/10.1007/bf03652036>

**Students' Evaluation of Information during Online Inquiry:
Working Individually or in Pairs**

Carita Kiili

Department of Education, University of Oslo, Norway
Current email: carita.kiili@tuni.fi

Julie Coiro

jcoiro@snet.net
School of Education, University of Rhode Island, USA

Eija Räikkönen

eija.m.raikkonen@jyu.fi
Faculty of Education and Psychology, University of Jyväskylä, Finland

Students' Evaluation of Information during Online Inquiry: Working Individually or in Pairs

Abstract

Varying information quality and an increase of misinformation on the Internet accentuates the importance of supporting students' competencies to critically evaluate information. This study compared how individuals and pairs of secondary students worked to evaluate the quality of online information across two inquiry topics. Two similar studies were conducted with 140 Finnish (Study I) and 52 US (Study II) students. Students were asked to conduct an online inquiry and then write an essay about one of two topics: allowing the genetic modification of organisms (GMO) or the effects of social media on people's quality of life (SM). Students worked either individually or in pairs. Their work was supported with a digital tool that prompted them to evaluate the credibility of online texts they selected as sources for their essays. Three separate analyses of covariance were conducted to examine differences in evaluation of online texts between individual and paired readers as well as between the inquiry topics across three dimensions: 1) total number of relevant justifications for credibility evaluations, 2) different types of relevant justifications, and 3) overall quality of students' evaluations and justifications across all responses captured by the digital tool. Results showed that working in pairs appeared to support the evaluation of online information in all assessed aspects in one context (Study I) but not in another (Study II). In Study 1, the GMO topic appeared to stimulate students' evaluation of information slightly more than SM, while there was no difference in evaluation performance across topics in Study 2. Findings suggest that discussing the credibility of online texts with a partner is a promising practice. Future research should explore more qualitative dimensions of how partners work together as they evaluate online texts and how instruction could be used to support collaborative evaluation.

Introduction

Readers assume much of the responsibility for evaluating the quality of online information due to the openness of the Internet (Metzger, 2007; Thomm & Bromme, 2012). Consequently, recent research (Bråten & Braasch, 2017) and national curricula (Australian Curriculum Assessment and Reporting Authority; National Governors Association Center for Best Practices & Council of Chief State School Officers, 2010) highlight the importance of supporting students' competencies to critically evaluate online information. In addition, an expert group set-up by the European Commission (2018) to counter disinformation online suggests five important pillars, one of which is the promotion of media and information literacy.

Furthermore, recent findings suggest more instruction is needed to develop students' ability to critically evaluate online information. School-age readers rarely stop to spontaneously consider the quality of information when engaging in online inquiry (Kiili, Laurinen, & Marttunen, 2008; Paul, Macedo-Rouet, Rouet, & Stadtler, 2017; Walraven, Brand-Gruwel, & Boshuizen, 2009). In addition, when students do stop to evaluate online texts, their judgments are often superficial and may lack any reasoned justification (Barzilai & Zohar, 2012; Coiro et al., 2015; Forzani, 2018). Finally, evaluation of biased or untrustworthy information seems to be particularly challenging for many readers (Kiili, Leu, Marttunen, Hautala, & Leppänen, 2018, Pérez et al., 2018; Tseng, 2018).

These results imply that students' attention needs to be directed with explicit prompts toward features that help them to make judgments about the quality of online texts (Bråten, McCrudden, Lund, Brante, & Strømsø, 2018; Paul, Stadtler, & Bromme, 2017). Other research suggests that peer-collaboration with opportunities to discuss different perspectives could potentially enhance deeper evaluation of online information (Dillenbourg, 1999; Teasley, 1995). Students, particularly those with limited knowledge of evaluation strategies, may benefit from pairing evaluation prompts with explicit models of effective strategies offered by their peers (Coiro, 2011; Nokes-Malach, Richey, & Gadgil,

2015).

In line with these two instructional practices, this study sought to clarify what kinds of justifications students offer for their credibility evaluations of texts during online inquiry when supported with prompts embedded in a digital tool, and whether working with a partner would foster evaluation of online information across two inquiry topics.

Critical Evaluation during Online Inquiry

Online inquiry refers to a set of complex, intertwined processes employed when individuals use the Internet to explore questions; search for, evaluate and synthesise information related to those questions; and communicate to others what was learned (Brand-Gruwel, Wopereis, & Vermetten, 2005; Leu, Kinzer, Coiro, Castek, & Henry, 2013). During these processes, skilled online readers build an understanding of the issue discussed across multiple online texts including information about sources, such as document type, author motives and author expertise (Bråten, Stadtler, & Salmerón, 2018; Perfetti, Rouet, & Britt, 1999). Making connections between a website's source and the online textual content allows readers to make reasoned judgments about the credibility, or perceived quality and believability, of the information encountered (Tseng & Fogg, 1999).

Evaluating the credibility of online text also plays an important role in how students learn during online inquiry (Bråten, Stømsø, & Britt, 2009; Wiley et al., 2009). Successful learners are more capable of discriminating reliable information from unreliable information (Wiley et al., 2009). In addition, the ability to reflect on the quality of selected online texts is common among readers who skilfully build their understanding across multiple texts (Cho, Woodward, Li, & Barlow, 2017). Furthermore, attention to credibility of online information appears to support adequate decision-making (Gerjets, Kammerer, & Werner, 2011) and argumentative source-based writing (Barzilai, Tzadok, & Eshet-Alkalai, 2015).

Scaffolds Supporting Evaluation of Online Texts

Metacognitive prompting, such as deep-level reasoning questions (e.g., Craig, Sullins, Witherspoon, & Gholson, 2006) and reflection prompts (e.g., van den Boom, Paas, van Merriënboer, & van Gog, 2004) have been shown to support learning processes. Similarly, students might benefit from digital or paper/pencil prompts that remind them to stop and consider the quality of online texts (Britt & Aglinskas, 2002; Gerjets et al., 2011; Kammerer, Meier, & Stahl, 2016).

To assist high school students' evaluation practices, Britt and Aglinskas (2002) developed computer-based scaffolds for exploring historical controversies. Prompts asked readers to identify, use, and evaluate documents, and to pay attention to source features and important content of each document. Prompts were designed to focus students' attention on author cues (e.g., name, position, author motives) and document cues (e.g., publication date, type of publication). Findings showed positive outcomes in identifying and evaluating documents when students used these scaffolds.

Prompts may be particularly useful for those students who are aware of strategies but do not naturally use them when reading online (Hogan & Vernhagen, 2012; Walraven et al., 2009). For example, Paul et al. (2017) interviewed 9th graders ($n = 44$) in relation to a brief multiple document reading task and found that students possessed source knowledge but utilised it only in certain conditions. Almost all students (96%) mentioned external prompting as an important factor that encouraged them to apply their sourcing knowledge. Thus, students who know specific evaluation strategies but do not necessarily apply them in practice may benefit from being prompted to execute them.

Collaboration Supporting Evaluation of Online Texts

In collaborative learning, two or more learners engage in coordinated, synchronous interaction aimed at achieving a common goal (Jeong & Hmelo-Silver, 2016; Roschelle & Teasley, 1995). In many cases, collaboration is beneficial for students' learning (Lou, Abrami,

& d'Apollonia, 2001; Springer, Stanne, & Donovan, 1999). A meta-analysis of 122 studies comparing small-group and individual learning with technology showed that small-group learning had a moderate positive effect on task performance and small positive effects on individual achievement on post-tests (Lou et al., 2001).

These positive results can be explained by the fact that peer-interaction provides opportunities for students to make their thinking explicit, negotiate their views, and build and extend one another's understanding (cf. Teasley, 1995; Dillenbourg, 1999). Further, students may also benefit from interacting with more capable learners who can help them to accomplish a task they could not accomplish alone (Vygotsky, 1986). Thus, by interacting with peers to solve a joint problem on the Internet, students can discuss the credibility of online texts and learn alternative strategies for successfully evaluating information during online inquiry (Coiro, Castek, & Guzniczak, 2011). Although working together offers mechanisms that often enhance learning, collaboration does not always result in better learning outcomes (Lou et al., 2001). Several associated factors may interfere with learning, including group dynamics, and lack of trust, sense of community or common goal (Kreijns, Kirschner, & Jochems, 2003).

Although some research exists concerning collaborative online inquiry (see Zhang & Quintana, 2012; Zhang, 2013), most research is small scale and qualitative in nature, zooming in on collaboration processes among small samples of students, but making it difficult to uncover more product-based outcomes across a larger sample (Castek, Coiro, Guzniczak, & Bradshaw, 2012; Kiili, Laurinen, Marttunen, & Leu, 2012; Knight & Mercer, 2017). In addition, few studies compare online reading performance among larger samples of individual and paired readers. Lazonder (2005) compared the search performance of nine pairs of readers and seven individuals and found that pairs used their time more effectively, used more versatile search strategies, and showed more monitoring and evaluation during

their searches. To our knowledge, no previous studies have compared individuals and pairs of readers' information evaluation during online inquiry. This study sought to clarify whether students benefitted from working with a partner when evaluating information during online inquiry with the support of digital scaffolds. Specific research questions included:

1. How did students evaluate the credibility of texts and justify their evaluations during online inquiry when supported by prompts embedded in a digital tool?
2. What was the overall quality of students' evaluations and justifications across all of their responses captured by the digital tool?
3. How did individual readers and pairs of readers perform when evaluating online texts across two different topics in terms of a) *total number* of relevant justifications, b) *different types* of relevant justifications, and c) *overall quality* of evaluations and justifications.

Methods

This paper presents the results of two similar studies, one conducted in Finland (Study I) and one in the United States (Study II). Researchers initially planned a quasi-experimental research design to represent students in both countries. However, due to several differences across the two countries (e.g. unequal sample sizes, varied student backgrounds, varied data collection procedures, and different course grades as control variables), it made more sense to answer our research questions using parallel, but separate, analyses.

Participants

Study I participants were 140 voluntary Finnish students (74 females; 66 males; $M = 16.7$ years; $SD = 0.62$) recruited from seven schools; 35 students were in their first year in upper-secondary school, and their ages varied from 15 to 17 years. The remaining students

($n = 105$) were in their second year of upper-secondary school, and their ages varied from 16 to 19 years.

Study I was integrated into a language arts class, and the completed task served as a final course exam. At the time of the exam, first-year students had taken two language arts courses, and second-year students had taken four of the six obligatory upper-secondary school courses. Eight voluntary Finnish language arts teachers participated in a half-day seminar to help plan the specific inquiry topics and data collection in their schools.

Study II was conducted approximately three months later and was designed to replicate procedures as much as possible. Participants were 52 Grade 10 US students (16 females; 36 males; $M = 15.7$ years; $SD = 0.85$) from two schools. Study II was integrated into six classes (four language arts and two history classes from one suburban high school in north-eastern US and two language arts classes from one urban high school in central US).

It was agreed that the US inquiry tasks would involve the same two topics suggested by the Finnish teachers. Consequently, while the completed task was separate from their regular class instruction, scores on the task were integrated into their final course grade as a classroom assignment. All participants had previously completed assignments in Grades 9 and 10 involving online research and argumentative writing as part of their district's initiatives to align with related educational standards.

Digital Tool Supporting Online Inquiry

Participants engaged in the task supported by the Online Inquiry Tool (Kiili, Coiro, & Hämäläinen, 2016). This tool was selected because it was explicitly designed to scaffold those aspects of online inquiry with which students seem to struggle when engaging in online inquiry (see Figure 1). The tool's purpose is to help students organise information from multiple online texts, evaluate online texts, analyse the argumentation of the texts, and regulate several complex cognitive activities likely to present challenges during online inquiry. Compared to other representational tools designed to support argumentative

discussion around controversial topics, this tool had additional features (space for URL-address and prompts to evaluate information credibility) that would direct students' attention to sources and their quality.

The present study focused only on what kinds of evaluative judgments the tool triggered during students' online inquiry in preparation for writing a source-based argumentation essay. The tool prompted readers to evaluate the credibility of each online text by selecting the most appropriate traffic light from three options: green indicated that the text appeared to be credible, yellow warranted some degree of caution, and red suggested that the text might not be credible. Then, students were asked to justify their evaluations by typing comments into a pop-up box that appeared after the appropriate traffic light was chosen.

---INSERT FIGURE 1 HERE---

Procedure

Training. Before engaging in the online inquiry task, students in both Study 1 and Study 2 were given a short training in a preceding lesson about how to use the Online Inquiry Tool. Teachers introduced and explained how to use different features of the tool in the context of an inquiry topic (different than the study topic). Teachers explained what each traffic light meant and introduced aspects, such as expertise, text purpose, and argument quality, which could be addressed when evaluating the credibility of web page information. However, no instruction was offered beyond highlighting these aspects as possible points for evaluation. After, students briefly practiced using the tool with two self-selected websites on the example topic.

Selecting inquiry topics. In a subsequent lesson, students conducted their online inquiry. All students were given a choice to conduct their inquiry about one of two topics: a) allowing the genetic modification of organisms (GMO) or b) the effects of social media (SM) on people's quality of life. This choice was intended to increase students' engagement

with the task (cf. Wigfield & Guthrie, 1997). The two given topics came from different disciplines, representing either science or social science, but both topics were controversial in nature and could be approached from different perspectives. In Study I, students chose the social media topic 54 times and GMO 34 times. In Study II, the corresponding numbers were 26 and 7.

Engaging in inquiry. Teachers assigned students to work either individually or in pairs to conduct online research on their selected topic (GMO or SM). They were asked to deliberate the topic from different perspectives by considering both supporting arguments and counter arguments of each perspective. To ensure comparability of their notes, students were asked to start their graph by inserting the given claim (i.e., “GMO should be allowed” or “Social media increases quality of life”). This procedure ensured that claims were clear, unambiguous, stated in the positive, and constant across readers of each topic. For the final essay, students were asked to take a position and justify their reasoning with evidence from their readings. Students were also given criteria for how their essays and their notes and evaluations in the graphic organizer (see Figure 1) would be assessed.

Across the two studies, there were some differences in how these procedures played out. In Study I, teachers ($n = 8$) chose the work mode that best served course learning objectives, resulting in 52 pairs and 36 individual readers. Students worked in two phases during one session. In Phase 1 (40–70 minutes), students explored online texts and filled in a graphic organiser. In Phase 2 (1.5–2.5 hours), students composed their essays while referring to their notes. Paired students worked together through all phases.

In Study II, teachers ($n = 6$) randomly assigned half of the students to work with assigned partners and the other half to work as individuals, resulting in 14 pairs and 19 individual readers. Students also worked in two phases, but due to scheduling restrictions, sessions took place over three consecutive days. Phase 1 (exploration of online texts) occurred over 1.5 sessions (70–80 minutes), and Phase 2 (essay writing) occurred over 1.5

sessions (60 minutes). Due to local school requirements, all students in Study II wrote their essays individually regardless of their working mode when exploring the online texts.

Data Analysis

Patterns in students' credibility evaluations and justifications. To answer question 1, we first examined how often students evaluated texts they used in their graphs as credible, somewhat, or not at all credible. Then, we identified relevant justifications and divided them into 16 different categories representing either authority or content (see Table 1). Categories emerged through content analysis (Krippendorff, 2004), first by deductively applying previously established criterion codes (Barzilai & Zohar, 2012; Kiili et al., 2008) and then by using inductive procedures (Bogdan & Biklen, 2003) that more accurately represented the set of evaluation practices observed in the data. The two first authors independently scored 30% of the justifications that were in English (Study II), reaching 87% agreement.

---INSERT TABLE 1 HERE---

Overall quality of credibility evaluations and justifications. To assess students' overall critical evaluation abilities as represented in the graphs (RQ2), we developed a six-point rubric reflecting students' ability to demonstrate versatility, depth, and some amount of scepticism across their credibility evaluations and justifications. When determining students overall critical evaluation ability, the unit of analysis was the whole graph. Data from students' evaluations (i.e., their traffic light selections) as well as from the previous analysis about their justifications informed our ratings. Table 2 shows the scoring criteria for the different ability levels represented in students' evaluations. Two researchers independently scored 33% of graphs from the English data (Study II), reaching 82% agreement.

---INSERT TABLE 2 HERE---

Statistical analysis of paired and individual work across two topics. To answer question 3, three *dependent variables* were first formed based on previous content analyses: 1) number of relevant justifications generated by students for their evaluations of online texts, 2) number of different types of relevant justifications (i.e., how many different justification types were represented), and 3) overall critical evaluation score. The *independent variables* were topic (0 = SM, 1 = GMO) and work mode (0 = pairs, 1 = individuals). In Study I (Finland), the *control variables* were report card grade in language arts and the number of language arts courses taken in upper secondary school (0 = four courses taken in the language arts, 1 = two courses taken in the language arts). In Study II (US), there was one control variable: report card grade in language arts.

Using SPSS software, univariate analysis of covariance (ANCOVA) was conducted to examine differences between individual readers and pairs of readers in terms of their performance in evaluating credibility of online texts (assessed by total number of relevant justifications, number of different types of justifications, and overall quality of evaluations in the argument graph). Three separate ANCOVAs were conducted in both studies due to the small sample sizes ($n = 88$ in Study I; $n = 33$ in Study II). In both studies, prior to analyses, the total number of justifications was normalised using logarithmic transformation. Additionally, report card grades were adjusted for in all analyses. As students in Study I were from two different language arts courses, the course level (second or fourth course during upper-secondary school) was adjusted for in the Study I analyses.

Results of Study I (Finland)

Descriptive Patterns in Credibility Evaluations and Their Justifications

With respect to question 1, Finnish students most often evaluated online texts that they included in the graph as credible (50.9% of texts). They expressed some caution towards the credibility of 37.1% of the texts but they seldom regarded texts that they included as not credible (7.8%). Students did not evaluate 4.2% of the online texts at all. The

majority of students (86.4%) indicated some scepticism (yellow or red light) in their evaluations.

Table 3 shows the frequencies of different types of justifications for students' credibility evaluations and the number of students (either individual readers or pairs of readers) who used each justification type at least once in their graph. Among all students, 88.6% (78 out of 88) included at least one relevant justification in their graphs.

---INSERT TABLE 3 HERE---

Content-focused justifications were more common (60.3% of all justifications) than authority-focused justifications (39.7%). Among the content-focused justifications, referring to the objectivity of information (or lack thereof) was the most common justification type, accounting for 20.1% of all relevant justifications. With respect to authority-focused justifications, students most often paid attention to the organization affiliated with an online text (19.6% of all justifications), with a little more than half of students using this type of justification at least once in their graphs. Notably, in spite of the fact that students were exploring a controversial issue and gathering both supporting and counter reasons from multiple texts in their graphs, they rarely used aspects of corroboration (1.0% of all justifications) in their evaluations.

Students' Abilities to Critically Evaluate Online Texts

With respect to question 2, nearly one third of Finnish students (29.6%) showed low ability to evaluate online texts that they selected for their graphs: 11.4% of students ($n = 10$) did not present any acceptable justifications in their graphs, 6.8% ($n = 6$) *sometimes* provided justifications but most were superficial or irrelevant, and 11.4% ($n = 10$) *usually* provided relevant justifications but they were mostly repetitive and superficial. Approximately one quarter of students in our sample ($n = 23$) scored in the proficient range, with evaluations and justifications that demonstrated some versatility but little scepticism. Finally, 34.1% of students ($n = 30$) demonstrated critical evaluation skills at the advanced levels and 10.2% (n

= 9) scored at the mastery level, with justifications demonstrating versatility, depth of reasoning, and some amount of scepticism.

Differences in Evaluation of Online Texts in Terms of Work Mode and Topic

With respect to question 3, there were no interaction effects between work mode and topic in Study 1 for any of three variables used to assess performance in evaluating online texts (p -values .12, .22, and .17, respectively). Instead, there were main effects for work mode for all three variables and main effects for topic for number of justifications and overall quality of evaluations (Table 4). Thus, working in pairs appeared to support the evaluation of online texts. Pairs presented more relevant justifications, their justifications were more diverse, and they scored higher on the overall quality of their evaluations compared with individual readers (Table 4, Figure 2). Effect sizes estimated using partial eta were large for the number of justifications ($\eta p^2 = .13$) and number of different justification types ($\eta p^2 = .17$) and medium for the overall quality of evaluations ($\eta p^2 = .06$).

--- INSERT TABLE 4 HERE---

--- INSERT FIGURE 2 HERE---

Results also showed that Finnish students who selected the GMO topic, whether working individually or in pairs, more actively justified their credibility evaluations compared with those who selected social media (Table 4, Figure 3). Additionally, both pairs and individuals showed higher overall quality when evaluating texts for GMO than for social media. As shown in Table 4, effect sizes for these two variables were either small or medium.

--- INSERT FIGURE 3 HERE---

Results of Study II (United States)

Credibility Evaluations and Their Justifications

With respect to question 1, US students often regarded the online texts included in their graphs as either credible (44.5%) or somewhat credible (34.5%). Only 1.7% of texts were

regarded as not credible. However, almost one fifth of the texts (19.3%) were not evaluated at all. In their evaluations, 69.7% of students indicated at least some scepticism.

In addition, 66.7% of US students (22 out of 33) presented at least one relevant justification when evaluating the credibility of online texts (Table 5). Content-based justifications accounted for 56.8% of all relevant justifications; thus, they were more common than authority-focused justifications. For content-focused justifications, objectivity (11.5% of all relevant justifications), and general impression or style (10.1%) were the more commonly used types of justifications. Among authority-focused justifications, attention to the website domain (e.g., “it’s a .com website”) was a frequently used justification, accounting for 22.3% of all relevant justifications.

---INSERT TABLE 5 HERE---

Students’ Abilities to Critically Evaluate Online Texts

For question 2, according to the rubric, close to half of the US students (45.5%) showed low ability to evaluate the online texts they selected for their graphs: 33.3% of students ($n = 11$) did not present any acceptable justifications in their graphs, 6.1% ($n = 2$) performed at an emerging level, and 6.1% ($n = 2$) at a developing level. Slightly more than one fourth of students (27.3%; $n = 9$) demonstrated overall proficiency; 21.2% ($n = 7$) reached advanced ability levels, and 6.1% ($n = 2$) demonstrated mastery levels of critical evaluation.

Differences in Evaluation of Online Texts in Terms of Work Mode and Topic

With respect to question 3, there were no interaction effects between work mode and topic in Study 2 for any of three evaluation variables (see Table 6). In addition, there were no main effects for work mode or topic. Results showed that whether US students worked in pairs or individually while researching either topic (GMO or SM), they presented similar numbers of justifications with similar levels of diversity, and they had similar scores on their overall ability to evaluate online texts.

--- INSERT TABLE 6 HERE---

Discussion

The present study sought to explore whether paired reading would support students' evaluation of online texts when prompted to do so during online inquiry, an area in which students have often been found to struggle. In this section, we discuss overall patterns in how students across the two samples evaluated online texts in terms of the justifications they provided (RQ1), and the overall versatility, depth, and scepticism demonstrated in their justifications (RQ2). Then, we discuss the extent to which collaboration may have supported students' evaluations (RQ3) and role that different topics may have played in the evaluation process (RQ3). We end our discussion by considering limitations and sharing some instructional ideas.

Students' Skills In Evaluating Online Texts

There were similar patterns, as well as differences, in the type, number, and quality of evaluations and justifications used by Finnish (Study I) and US (Study II) students. Across both Finnish and US samples, students expressed more content-focused justifications than authority-focused justifications with references to objectivity, or lack thereof, being the most frequent content-focused justification in both studies. This pattern is consistent with other studies (e.g., Coiro et al., 2015; Goldman, Braasch, Wiley, Graesser, & Brodowinska, 2012) that reported readers tend to focus more on relevance and the quality of information and less on the authors themselves.

It may be that particular features of our online inquiry tool facilitated these content-focused justifications. That is, our digital tool was designed to provide a separate space for readers to insert relevant content-based claims and explicit prompts in the justification box to direct students' to rate the quality of those claims. While this possibility needs to be explored more systematically in future studies, it may be that the visual prompts in our graph reminded readers to balance their attention to author credibility with the need to judge the quality of semantic content provided by that author; two important sourcing

strategies recommended when reading multiple documents (see Braasch, Bråten, & McCrudden, 2018).

When we looked more closely across patterns of authority-focused justifications, findings varied. Among all Finnish students, the most common justification type was related to the organization affiliated with the web page (19.6% of all justifications). In Study II, the most common justification type was appealing to the website domain (.org; .com). Among all US students, 40% used this justification, and it accounted for 22.3% of all justifications. Finnish students did not use this justification type at all.

In the US, many librarians and teachers emphasise the inferences that can be made from a URL address (e.g., com; gov) - see for example <http://guides.lib.uw.edu/research/evaluate/domains>. In contrast, Finnish websites – government, education, and commercial – can all have the domain .fi and, therefore, not as many inferences can be made based on the domain. It is important that students pay attention to the domain and make inferences about the authority of the website. However, if students systematically rely only on domain information, their evaluations will remain quite superficial. Thus, while the domain may be useful for making initial evaluations, instruction should emphasize how to more carefully explore information about a webpage's author and content to obtain a deeper understanding of its quality.

Findings showed that many students in both studies (29.6% in the Finnish sample and 45.5% in the US sample) demonstrated low levels of ability to evaluate texts during online inquiry. Only 10.2% (Finnish sample) and 6.1% (US sample) reached the mastery level, with justifications demonstrating versatility, depth of reasoning, and some amount of scepticism. Despite these similar trends, some of the performance differences across the two samples may be explained by the higher average age of students in the Finnish sample ($M = 16.7$) compared to those in the US sample ($M = 15.7$). Older, more mature, students who

have an additional year of language arts instruction are likely to have more skills related to judging the quality of information in various contexts.

Overall, however, the low percentages of students across both Finnish and US samples who demonstrated mastery in evaluating the quality of online information points to a need for increased attention to further developing students' critical evaluation practices. In the present studies, only very general prompts were used. Students were simply asked to evaluate the credibility of one online text at a time and to justify their evaluation without giving any hints about the aspects of credibility they could attend to. Future studies could explore whether more specific prompts that ask students to evaluate quality of argumentation, use of evidence, and text purpose, would foster deeper levels of evaluation. Students may, for example, be prompted to first evaluate the collection of texts individually before discussing with their partner. This might focus students' attention on opportunities to corroborate one source with another, an important skill that was rarely observed in either the Finnish or US sample.

The Role of Collaboration and Topic In Critical Evaluation

Role of collaboration. Regarding findings about whether individual or paired reading during online inquiry facilitates more advanced practices for evaluating online texts, results were inconsistent. While US students seemed to work equally well regardless of their work mode, students in the Finnish sample seemed to benefit from working in pairs when prompted to evaluate the quality of texts during online inquiry. These findings coincide with work emerging elsewhere (Lou et al., 2001) suggesting that each reading situation involves unique conditions related to readers, contexts, and materials (Bråten et al., 2018; Britt, Rouet, & Durik, 2017) and that the act of partnering students, by itself, does not necessarily ensure productive collaboration (Häkkinen & Mäkitalo-Siegl, 2007; King & Rosenshine, 1993).

When looking more closely at the differences between pairs and individuals in the Finnish sample, pairs presented more justifications for their evaluations than individual readers did, and the pairs' justifications were more versatile, demonstrating a large effect for amount and versatility. The better performance of paired readers might be explained by opportunities to make ones' thinking explicit and negotiate their credibility evaluations (Dillenbourg & Schneider, 1995). In the current study, having a partner with whom to discuss the most appropriate credibility rating for each online text may have promoted more opportunities to jointly negotiate their justifications. Students may also bring different strategic resources to the learning situation (cf. Nokes-Malach et al., 2015) or ask thought provoking questions (King, 2002) that could partly explain the richer repertoire of strategies found in the pairs' graphs.

Similarly, with respect to the overall quality of evaluations and justifications in the graphs, paired readers in the Finnish sample scored higher than individual readers. However, differences in overall quality demonstrated only a medium effect. Thus, even though students were able to apply a variety of different justification types in their evaluations, the quality of evaluations did not, in general, reach a very high level. This is in line with previous results showing the same tendency (e.g., Coiro, Coscarelli, Maykel, & Forzani, 2015; Johnson, Sbaffi, & Rowley, 2016). Future research should concentrate on providing students with discussion supports that facilitate the quality of discourse around online evaluation processes to increase the likelihood that partner work can foster more advanced evaluation products.

However, previous studies have shown that transfer effects are difficult to achieve (Walraven, Brand-Gruwel, & Boshuizen, 2013) and might require longer interventions combined with modeling (Coiro, 2011), class discussions (Applebee, Langer, Nystrand, & Gamoran, 2003), and timely feedback (Nicol & Macfarlane-Dick, 2006). Teacher-led discussions during or after critical evaluation tasks might increase the likelihood that all

students, including pairs demonstrating shallow evaluation behaviours, are introduced to a wider variety of evaluation strategies.

Role of inquiry topics. In both studies, students favoured social media (SM) over GMO. In Study I, whether working individually or in pairs, inquiry focused on GMO stimulated students to more actively justify their evaluations, which in turn resulted in a higher overall quality of evaluations compared with those of students who selected the social media task. These effects were, however, smaller than the effects of work mode. These differences might be explained by topic familiarity (see also Bråten et al., 2018), since social media was probably more familiar to students in our sample, and judgments on texts might not have necessitated as much explanation due to this familiarity. There might have also been differences in the types of the texts that students selected, which could have stimulated different evaluation strategies. However, we did not explore these possible explanations in the present study. In Study II, the topic did not have any effect on students' evaluation of texts. This might be due to smaller sample size and the fact that only a very few students chose GMO as their inquiry topic.

Limitations

Many previous studies have investigated students' evaluation skills when using a limited amount of texts (Barzilai et al., 2015; Wiley et al., 2009). Yet, one strength of the present study was that it explored students' evaluation of online texts in a completely unconstrained Internet space. In addition, the present study involved an inquiry task designed in close collaboration with teachers that enabled students to choose from two inquiry topics. Nevertheless, this open and authentic design also introduced several limitations, including an unequal number of students in the different groups (work mode and topic), the possibility of a teacher effect, and the incorporation of a wide range of different texts into different students' graphs, making it difficult to compare consistent strategy use.

Further, as the studies were integrated into typical classroom routines in each country, district expectations (whether or not individuals could complete an assessment with a partner), grading differences (final course exam vs. class assignment), and class times (one longer session vs. three shorter sessions) varied across the two contexts, making it difficult to follow the same research protocols in both countries. Consequently, we were unable to truly compare the results statistically across two samples from different countries. To better understand how partner work impacts online inquiry and the possibility of cultural differences influencing the quality of evaluation products (i.e., credibility justifications, argumentation essays), there is a need to develop valid and reliable ways of capturing evidence of critical evaluation skills in an open Internet context while considering all of these design complexities.

Instructional Implications

Our findings suggest that dyadic face-to-face discussions can, under some conditions, facilitate more active, versatile, and deeper evaluation of online information compared to individual work. Future research should examine the potential impact on final justified credibility judgements with more extensive instruction focused on discussing and taking notes on different aspects of information credibility with a partner.

Because we found that some students, even when working with a partner, did not evaluate information at all, it seems important to find ways of motivating students to actively engage in critical evaluation practices (see Paul et al., 2017). The importance of evaluation could be underscored by discussing, for example, the potentially negative consequences of uncritical inspection of online information in daily life (see, for example, McGrew, Ortega, Breakstone, & Wineburg, 2017).

Further, our study suggests that students (individuals and pairs) need additional support to reach advanced levels of reasoning around credibility judgements. For example, we seldom observed students' efforts to evaluate quality of argumentation. Deeper

reasoning could be supported with more explicit modelling of how to use multiple evaluation criteria (see Coiro, 2011) or specific prompts asking students to reflect on the accuracy of different claims and related evidence. Questions that ask students to observe flaws in scientific claims may also provide practice in becoming more critical collaborative readers (Tseng, 2018).

Finally, because critical evaluation is a complex skill, students would likely benefit from working with different partners across the school year. This would provide students opportunities to work with peers with different proficiency levels, motivations, and strategic repertoires as they jointly reasoned about the credibility of online information. Over time, these practices could become part of the long-term, systematic instruction needed to develop the complex skills associated with critical evaluation.

Acknowledgements

The research reported in this article was supported by the Elva Knight award 2014 from International Literacy Association to Julie Coiro and Carita Kiili. The authors want to thank Janna Inkeroinen, for helping in the data analyses.

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, 685–730. <https://doi.org/10.3102/00028312040003685>
- Australian Curriculum Assessment and Reporting Authority. (n.d.). *The Australian Curriculum, v6.0*. Retrieved from www.australiancurriculum.edu.au/Home
- 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>
- Barzilai, S. & Zohar, A. (2012). Epistemic thinking in action: Evaluating and integrating

- online sources. *Cognition and Instruction*, 30, 39–85.
<https://doi.org/10.1080/07370008.2011.636495>
- Bogdan, R. C., & Biklen, S. K. (2003). *Qualitative research for education: An introduction to theories and methods (4th ed.)*. Boston, MA: Allyn & Bacon.
- Brand-Gruwel, S., Wopereis, I., & Vermetten, Y. (2005). Information problem solving by experts and novices: Analysis of a complex cognitive skill. *Computers in Human Behavior*, 21, 487–508. <https://doi.org/10.1016/j.chb.2004.10.005>
- Bråten, I., & Braasch, J. L. G. (2017). Key issues on students' critical reading and learning in the 21st century information society. In C. Ng & B. Bartlett (Eds.), *Improving reading and reading engagement in the 21st century: International research and innovations* (pp. 77–98). Singapore: Springer.
- 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, M. A. Britt, & D. N. Rapp (Eds.), *Handbook of discourse processes* (2nd. ed., pp. 141–166). New York, NY: Routledge.
- Bråten, I., Strømsø, H. I., & Britt, M. A. (2009). Trust matters: Examining the role of source evaluation in students' construction of meaning within and across multiple texts. *Reading Research Quarterly*, 44, 6–28. <https://doi.org/10.1598/RRQ.44.1.1>
- Britt, M. A., & Aglinskias, C. (2002). Improving students' ability to identify and use source information. *Cognition and Instruction*, 20, 485–522.
https://doi.org/10.1207/S1532690XCI2004_2
- Britt, M. A., Rouet, J. F., & Durik, A. M. (2017). *Literacy beyond text comprehension: A theory of purposeful reading*. New York, NY: Routledge.

- Castek, J., Coiro, J., Guzniczak, L., & Bradshaw, C. (2012). Examining peer collaboration in online inquiry. *The Educational Forum*, 76, 479–496.
<https://doi.org/10.1080/00131725.2012.707756>
- Cho, B.-Y., Woodward, L., Li, D., Barlow, W. (2017). Examining adolescents' strategic processing during online reading with a question-generating task. *American Educational Research Journal*, 54, 691–724. doi:10.3102/0002831217701694
- Coiro, J. (2011). Talking about reading as thinking: Modeling the hidden complexities of online reading comprehension. *Theory Into Practice*, 50, 107–115.
<https://doi.org/10.1080/00405841.2011.558435>
- Coiro, J., Castek, J., & Guzniczak, L. (2011). Uncovering online reading comprehension processes: Two adolescents reading independently and collaboratively on the Internet. *60th Yearbook of the Literacy Research Association*, 354–369.
- 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>
- Craig, S. D., Sullins, J., Witherspoon, A., & Gholson, B. (2006). The deep-level reasoning questions effect: The role of dialogue and deep-level reasoning questions during vicarious learning. *Cognition and Instruction*, 24, 565–591.
https://doi.org/10.1207/s1532690xci2404_4
- Dillenbourg P. (1999). What do you mean by collaborative learning? In P. Dillenbourg (Ed.), *Collaborative-learning: Cognitive and Computational Approaches* (pp. 1–19). Oxford: Elsevier.
- Dillenbourg, P., & Schneider D. (1995). Mediating the mechanisms which make collaborative learning sometimes effective. *International Journal of Educational Telecommunications*, 1 (2–3), 131–146.

- European Commission, Directorate-General for Communication Networks, Content and Technology. (2018). *A multi-dimensional approach to disinformation: report of the independent high level group on fake news and online disinformation*. Brussels: European Union. doi:10.2759/739290 Retrieved from <https://www.cato.org/publications/policy-analysis/risky-business-role-arms-sales-us-foreign-policy>
- 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>
- 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>
- Goldman, S. R., Braasch, J. L. G., Wiley, J., Graesser, A. C., & Brodowinska, K. (2012). Comprehending and learning from internet sources: Processing patterns of better and poorer learners. *Reading Research Quarterly*, 47, 356–381. <https://doi.org/10.1002/RRQ.027>
- Hogan, N., & Vernhagen, C. (2012). Critical appraisal of information on the Web in practice: Undergraduate students' knowledge, reported use, and behaviour. *Canadian Journal of Learning and Technology*, 38(1), 1–14. <http://dx.doi.org/10.21432/T23K5P>
- Häkkinen, P., & Mäkitalo-Siegl, K. (2007). Educational perspectives on scripting CSCL. In F. Fischer, I. Kollar, H. Mandl, & J. M. Haake (Eds.), *Scripting computer-supported collaborative learning: Cognitive, computational and educational approaches* (pp. 263–271). New York, NY: Springer.
- 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>
- Johnson, F., Scaffi, L., & Rowley, J. (2016). Students' approaches to the evaluation of digital information: Insights from their trust judgments. *British Journal of Educational Technology*, 47, 1243–1258. <https://doi.org/10.1111/bjet.12306>
- 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., Coiro, J., & Hämäläinen, J. (2016). An online inquiry tool to support the exploration of controversial issues on the Internet. *Journal of Literacy and Technology*, 17(1–2), 31–52.
- Kiili, C., Laurinen, L., & Marttunen, M. (2008). Students evaluating Internet sources – From versatile evaluators to uncritical readers. *Journal of Educational Computing Research*, 39, 75–95. <https://doi.org/10.2190/EC.39.1.e>
- Kiili, C., Laurinen, L., Marttunen, M., & Leu, D. J. (2012). Working on understanding during collaborative online reading. *Journal of Literacy Research*, 44, 448–483.
<https://doi.org/10.1177/1086296X12457166>
- 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>
- King, A. (2002). Structuring peer interaction to promote high-level cognitive processing. *Theory into Practice*, 41, 33–39.
- King, A., & Rosenshine, B. (1993). Effect of guided cooperative questioning on children's knowledge construction. *Journal of Experimental Education*, 61, 127–148.
<https://doi.org/10.1080/00220973.1993.9943857>

- Knight, S., & Mercer, N. (2017). Collaborative epistemic discourse in classroom information-seeking tasks. *Technology, Pedagogy and Education, 26*, 33–50.
<https://doi.org/10.1080/1475939X.2016.1159978>
- Kreijns, K., Kirschner, P. A., & Jochems, W. (2003). Identifying the pitfalls for social interaction in computer-supported collaborative learning environments: A review of the research. *Computers in Human Behaviour, 19*, 335–353.
[https://doi.org/10.1016/S0747-5632\(02\)00057-2](https://doi.org/10.1016/S0747-5632(02)00057-2)
- Krippendorff, K. (2004). *Content analysis: An introduction to its methodology*. Sage.
- Lazonder, A. W. (2005). Do two heads search better than one? Effects of student collaboration on web search behaviour and search outcomes. *British Journal of Educational Technology, 36*, 465–475. <https://doi.org/10.1111/j.1467-8535.2005.00478.x>
- Leu, D. J., Kinzer, C. K., Coiro, J., Castek, J., & Henry, L. A. (2013). New literacies: A dual level theory of the changing nature of literacy, instruction, and assessment. In D.E. Alvermann, N.J. Unrau & R.B. Ruddell (Eds.), *Theoretical models and processes of reading* (6th ed., pp. 1150–1181). Newark, DE: International Reading Association.
- Lou, Y., Abrami, P. C., & d'Apollonia, S. (2001). Small group and individual learning with technology: A meta-analysis. *Review of Educational Research, 71*, 449–521.
<https://doi.org/10.3102/00346543071003449>
- McCrudden, M. T., Bråten, I., & Braasch, J. L. (2018). Introduction to research on multiple source use. In J. L.G. Braasch, I. Bråten & M.T. McCrudden (Eds.), *Handbook of multiple source use* (pp. 1–13). New York, NY: 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*, 147–162. <http://dx.doi.org/10.1037/edu0000057>

- McGrew, S., Ortega, T., Breakstone, J., & Wineburg, S. (2017). The challenge that's bigger than fake news: Civic reasoning in a social media environment. *American Educator*, Fall Issue, 4-7.
- Metzger, M. J. (2007). Making sense of credibility on the Web: Models for evaluating online information and recommendations for future research. *Journal of the American Society for Information Science and Technology*, 58, 2078-2091. <https://doi.org/10.1002/asi.20672>
- National Governors Association Center for Best Practices & Council of Chief State School Officers. (2010). *Common Core State Standards for English language arts and literacy in history/social studies, science, and technical subjects*. Washington, DC.
- 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, 199-218. <https://doi.org/10.1080/03075070600572090>
- Nokes-Malach, T. J., Richey, J. E., & Gadgil, S. (2015). When is it better to learn together? Insights from research on collaborative learning. *Educational Psychology Review*, 27, 645-656. <https://doi.org/10.1007/s10648-015-931>
- 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>
- Paul, J., Stadtler, M., & Bromme, R. (2017). Effects of a sourcing prompt and conflicts in reading materials on elementary students' use of source information. *Discourse Processes*. Advance online publication. <https://doi.org/10.1080/0163853X.2017.1402165>
- 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.
- 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.
- Roschelle, J., & Teasley, S. (1995). The construction of shared knowledge in collaborative problem solving. In C. O'Malley (Ed.), *Computer-supported collaborative learning* (pp. 69–197). Berlin, Germany: Springer Verlag.
- Springer, L., Stanne, M. E., & Donovan, S. S. (1999). Effects of small-group learning on undergraduates in science, mathematics, engineering, and technology: A meta-analysis. *Review of Educational Research, 69*, 21–51.
<https://doi.org/10.3102/00346543069001021>
- Teasley, S. D. (1995). The role of talk in children's peer collaborations. *Developmental Psychology, 31*, 207–220. <https://doi.org/10.1037/0012-1649.31.2.207>
- Thomm, E., & Bromme, R. (2012). "It should at least seem scientific!" Textual features of "scientificness" and their impact on lay assessments of online information. *Science Education, 96*, 187–211. <https://doi.org/10.1002/sce.20480>
- Tseng, A. S. (2018). Students and evaluation of web-based misinformation about vaccination: critical reading or passive acceptance of claims? Advanced online publication. *International Journal of Science Education, Part B*.
<https://doi.org/10.1080/21548455.2018.1479800>
- Tseng, S., & Fogg, B. J. (1999). Credibility and computing technology. *Communications of the ACM, 42*(5), 39–44. <https://doi.org/10.1145/301353.301402>
- van den Boom, G., Paas, F., van Merriënboer, J. J. G., & van Gog, T. (2004). Reflection prompts and tutor feedback in a web-based learning environment: Effects on students'

- self-regulated learning competence. *Computers in Human Behavior*, 20, 551–567.
<https://doi.org/10.1016/j.chb.2003.10.001>
- Vygotsky, L. S. (1986). *Thought and language*. Cambridge, MA: MIT Press.
- 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, 234–246. <https://doi.org/10.1016/j.compedu.2008.08.003>
- Walraven, A., Brand-Gruwel, S., & Boshuizen, H. P. (2013). Fostering students' evaluation behaviour while searching the Internet. *Instructional Science*, 41, 125–146.
<https://doi.org/10.1007/s11251-012-9221-x>
- Wigfield, A., & Guthrie, J. T. (1997). Relations of children's motivation for reading to the amount and breadth of their reading. *Journal of Educational Psychology*, 89, 420–432.
<http://dx.doi.org/10.1037/0022-0663.89.3.420>
- Wiley, J., Goldman, S. R., Graesser, A. C., Sanchez, C. A., Ash, I. K., & Hemmerich, J. A. (2009). Source evaluation, comprehension, and learning in Internet science inquiry tasks. *American Educational Research Journal*, 46, 1060–1106.
<https://doi.org/10.3102/0002831209333183>
- Zhang, M. (2013). Supporting middle school students' online reading of scientific resources: Moving beyond cursory, fragmented, and opportunistic reading. *Journal of Computer Assisted Learning*, 29, 138–152. <https://doi.org/10.1111/j.1365-2729.2012.00478.x>
- Zhang, M., & Quintana, C. (2012). Scaffolding strategies for supporting middle school students' online inquiry processes. *Computers & Education*, 58, 181–196.
<https://doi.org/10.1016/j.compedu.2011.07.016>

Carita Kiili is a postdoctoral fellow at the Department of Education, University of Oslo, Norway. Her research interests lie on digital literacies, especially on online inquiry and reading and composition of multimodal texts. She can be reached via email at c.p.s.kiili@iped.uio.no

Julie Coiro is an associate professor in the School of Education at the University of Rhode Island, USA. She teaches undergraduate and graduate courses in reading and co-directs the Graduate Certificate in Digital Literacy and Phd in Education program. She can be reached via email at jcoiro@snet.net

Eija Räikkönen is a senior lecturer at the Faculty of Education and Psychology, University of Jyväskylä, Finland. Her research and teaching interests include psychometrics. She can be reached via email at eija.m.raikkonen@jyu.fi

The interface is titled "Online Inquiry Tool". At the top left, there is a menu with options: "Create new chart", "Append chart", "Open chart", "Save chart", "Palette of Perspectives", and "Instructions". Below the menu is a "Claim" section with a text area labeled "<Add claim here>". The main workspace is divided into four columns: "Perspective", "Reasons for the claim", "Reasons against the claim", and "Synthesis". Each column has a text area for input and a traffic light icon (red, yellow, green) for status. The "Reasons for the claim" and "Reasons against the claim" columns have buttons for "Add reason for the claim" and "Add reason against the claim" respectively. A green bar at the bottom of the main area says "Add perspective". Below the main area is a "Notes" section with a text area for writing notes. A pop-up box labeled "Why do you think so?" is overlaid on the "Reasons against the claim" column, containing a text area for justifications and an "OK" button. Number 1 points to the traffic light icons, and Number 2 points to the pop-up box.

Figure 1. Online Inquiry Tool. Number 1 refers to the traffic lights and Number 2 to the pop-up box for students' justifications.

COLLABORATIVE EVALUATION

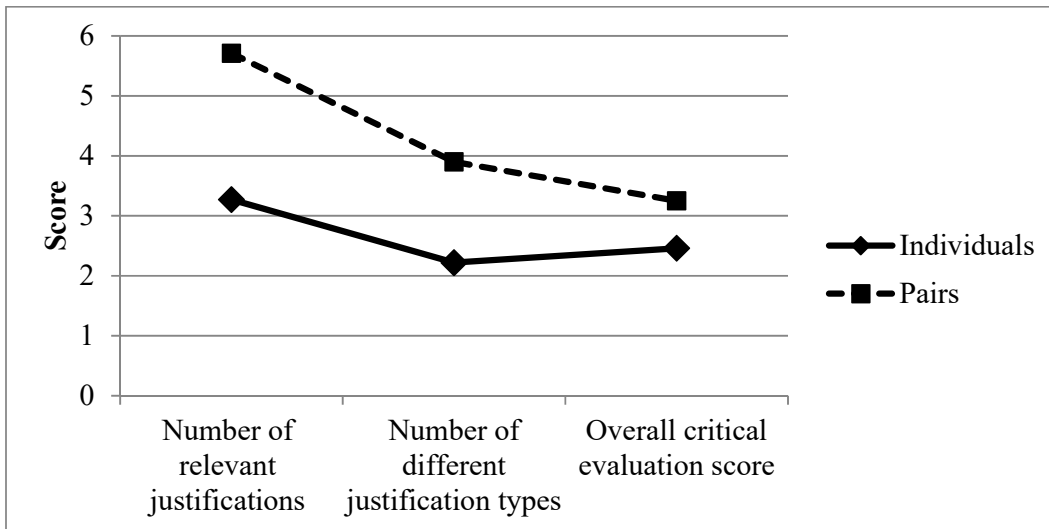


Figure 2. Evaluation of online texts according to work mode in Study I.

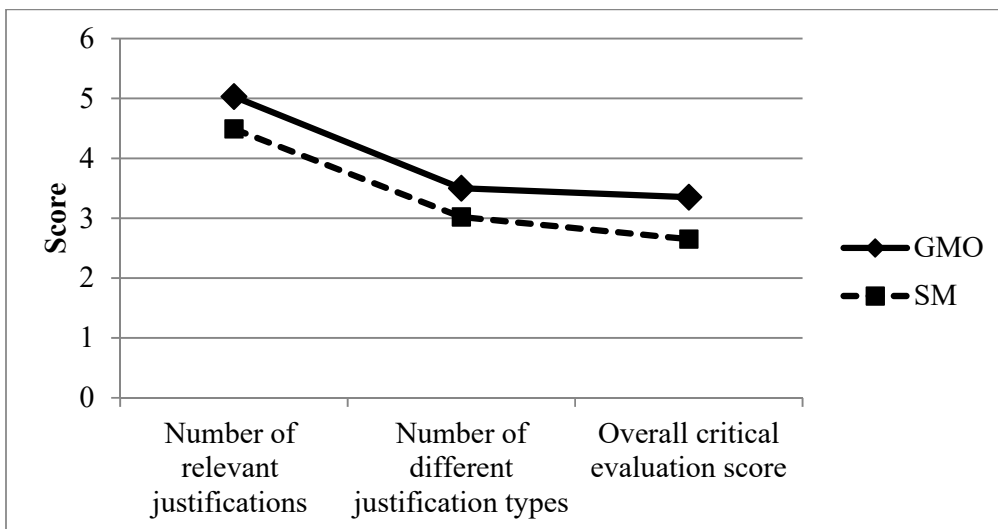


Figure 3. Evaluation of online texts according to task topic in Study I.

Table 1

Justification types for evaluations of online texts

<i>Justification type</i>	<i>Description</i>	<i>An example</i>
Authority-focused justifications		
Practices of the Internet forum	Evaluation is justified by considering the courses of actions of the Internet forum, such as Wikipedia or discussion forum. Justifications might concern, for example, how the quality of information has been controlled.	Usually, there is credible information on Wikipedia, and there are always sources available. If the information is not certain or the source has not been added, it is said (ID 114, individual reader).
Author of the text or the person interviewed	Evaluation is justified by focusing on attributes of the person who has written the text or who has been interviewed. Justifications can concern societal position, position in organizations, status, education, reputation, or expertise.	The authors are researchers on information and communication technology and they work at the University of Jyväskylä and University of Tampere (ID 294, pair of readers).
Organization affiliated with the Web page	Evaluation is justified by focusing on organization affiliated with the Web page, such as a university, political party, enterprise, newspaper, or magazine. Justifications might concern the expertise, authoritativeness, or reputation of the organization.	The Academy of Finland is devoted to high-quality research (ID 269, pair of readers).
Domain	Evaluation is justified by referring to the domain of the website.	It is .org, which means organization (ID 426, pairs of readers).
Author or publisher information provided	Evaluation is justified by paying attention to whether the Web page provides author or publisher information.	The author can be contacted (ID 117, individual reader).
Content-focused justifications		
Use of sources	Evaluation is justified by noting the presence or absence of references.	On the website, there were references to other scientific sources (ID 431, individual reader).

Intention towards good purposes	Evaluation is justified by noting that the Web page is created to inform readers, such as an educational page.	The website is created for educational purposes (ID 129, individual reader).
Objectivity	Evaluation is justified by commenting on the objective, biased, or purpose-oriented mode of the writing, or the representativeness of the opinion.	It was a blog and it only presented the views of one person (ID 260, pairs of readers).
General impression and style	Evaluation is justified by the appearance or design of the Web page.	The design of the website is convincing (ID 271 pair of readers).
Informativeness	Evaluation is justified by its informativeness, referring to the amount of information, the expository nature of the text, or the content coverage.	They give a lot of information about media education (ID 265, pairs of readers).
Quality of argumentation	Evaluation is justified by attributes related to the quality of argumentation, such as its balance (i.e., including both supporting and counter arguments), scope (text includes several perspectives), or lack of quality evidence.	There is research behind the article, but there are only a few arguments (ID 261, pairs of readers).
Correspondence with one's previous knowledge	Evaluation is justified by comparing information on the Web page with one's previous knowledge.	The website provides information that I already knew beforehand (ID 256, pair of readers).
Corroboration	Evaluation is justified by comparing and corroborating information provided on the Web page with information provided by one, or several, other Web resources.	There is similar information from multiple sources (ID 311, individual reader).
Research-basis of information	Evaluation is justified by the idea that information is either based on research or not.	More than one study is mentioned, and one of them also mentions the name of the researcher, but it does not provide a link to the research (ID 122, individual reader).

Currency	Evaluation is justified by the currency of information or by the observation that information is not that current.	This has been published in 2014 and, thus, it is current (ID 288, pairs of readers).
Other	Evaluation is justified by a relevant justification that does not fit any other category.	The website is provided in multiple languages, which increases its credibility (ID 244, pairs of readers).

Table 2

Scoring rubric for overall quality of students' evaluations and justifications in their graphs

Score	Ability level	Criteria
0	Not present or unacceptable	The graph does not include any evaluations and/or evaluations are not justified.
1	Emerging	<p>Only some (less than 75% of evaluations) of the evaluations are justified.</p> <p>Most justifications are repetitive and/or superficial (e.g., appearance, URL, or short statements) or just not relevant.</p>
2	Developing	<p>Most (75% or more) evaluations are justified.</p> <p>Justifications are repetitive and/or superficial or just not relevant.</p>
3	Proficient	<p>The graph includes at least two online texts. Most (75% or more) evaluations are justified.</p> <p>Justifications are somewhat versatile (relying on either author or content but not both) but mostly superficial.</p> <p>Little or no element of doubt is reflected in evaluations.</p>
4	Advanced	<p>The graph includes at least two online texts. Most (75% or more) evaluations are justified.</p> <p>Most justifications for evaluations are versatile (relying on author and content) and some reflect deeper levels of reasoning. Some element of skepticism is present.</p> <p>If a online text is evaluated as somewhat reliable, only one of the following aspects is considered: what aspect makes the information credible or what aspect weakens the credibility.</p>
5	Mastery	<p>The graph includes at least two online texts. Most (75% or more) evaluations are justified.</p> <p>Most justifications for evaluations are versatile (relying on both author and content) and reflect deeper levels of reasoning. Some element of skepticism is present.</p> <p>If the online text is evaluated as somewhat reliable, most justifications consider both what aspect makes the information credible and what aspect weakens the credibility.</p>

Table 3

Study I based on the Finnish data: number of justifications for evaluations by justification type among individual readers and pairs of readers (n = 88) and number of students (either individual readers or pairs of readers) who used the justification type at least once in their graphs.

Justification category	f	%	Number of students	%
Relevant justifications				
Content-focused justifications				
Objectivity	84	20.10	47	66.83
Research-basis of information	35	8.37	22	25.00
Use of sources	27	6.46	19	21.59
General impression or style	25	5.98	18	20.45
Quality of argumentation	22	5.26	18	20.45
Informativeness	22	5.26	19	21.59
Currency	12	2.87	6	6.82
Intention to good purposes	9	2.15	9	10.23
Correspondence with one's prior knowledge	7	1.67	6	6.82
Other	5	1.20	5	5.68
Corroboration	4	0.96	4	4.55
Content-focused justifications total	252	60.29	72	81.82
Authority-focused justifications				
Organization affiliated with the Web page	82	19.62	45	51.14
Author of the text or person interviewed	46	11.00	36	40.91
Practices of the Internet forum	22	5.26	19	21.59
Author or publisher information provided	16	3.83	12	13.64
Domain (URL address)	0	0	0	0
Authority-focused justifications total	166	39.71	61	69.32
Relevant justifications total	418	100	78	88.64

Table 4

Study I based on the Finnish data: means (M), standard deviation (SD) and results of ANCOVAs with main effects of work mode and topic in relation to three aspects of the evaluation of online texts. Report card grade and level of the language arts class were controlled for in the analyses

		<i>M</i>	<i>SD</i>	<i>F</i> (1, 83)	<i>p</i>	η_p^2
Work mode						
Number of relevant justifications* (range 0–14)	Individuals	3.27	2.64	12.13	<0.001	.13
	Pairs	5.71	3.26			
Number of different justification types (range 0–8)	Individuals	2.22	1.72	16.55	<0.001	.17
	Pairs	3.90	1.85			
Overall critical evaluation score (range 0–5)	Individuals	2.46	1.76	5.48	.02	.06
	Pairs	3.25	1.20			
Topic						
Number of relevant justifications* (range 0–14)	GMO	5.03	2.39	4.48	.04	.05
	Social Media	4.49	3.67			
Number of different justification types (range 0–8)	GMO	3.50	1.66	2.91	.09	.03
	Social Media	3.02	2.14			
Overall critical evaluation score (range 0–5)	GMO	3.35	1.23	4.85	.03	.06
	Social Media	2.65	1.60			

* For the analysis, variable was normalized, but original *Ms* and *SDs* are reported here.

Table 5

Study II based on the US data: number of justifications for evaluations by justification types among individual readers and pairs of readers (n = 33) and number of students (either individual readers or pairs of readers) who used the justification type at least once in their graphs.

Justification type	f	%	Number of students	%
Relevant justifications				
Content-focused justifications				
Objectivity	17	11.49	7	21.21
General impression or style	15	10.14	10	30.30
Research-basis of information	13	8.78	5	15.15
Informativeness	13	8.78	6	18.18
Other	9	6.08	5	15.15
Currency	6	4.05	3	9.09
Intention to good purposes	4	2.70	4	12.12
Quality of argumentation	4	2.70	3	9.09
Corroboration	2	1.35	2	6.06
Use of sources	1	0.68	1	3.03
Correspondence with one's prior knowledge	0	0	0	0
Content-focused justifications total	84	56.75	20	60.61
Authority-focused justifications				
Domain (URL address)	33	22.30	13	39.39
Organization affiliated with the Web page	17	11.49	8	24.24
Author of the text or person interviewed	12	8.11	8	24.24
Practices of the Internet forum	2	1.35	1	3.03
Author or publisher information provided	0	0	0	0
Authority-focused justifications total	64	43.25	18	54.55
Relevant justifications total	148	100	22	66.67

Table 6

Study II based on the US data: means (M), standard deviation (SD,) and results of ANCOVAs with main effects of work mode and topic in relation to three aspects of evaluation of online texts. Report card grade was controlled for in the analyses.

		<i>M</i>	<i>SD</i>	<i>F</i> (1, 28)	<i>p</i>	η_p^2
Work mode						
Number of relevant justifications* (range 0-20)	Individuals	3.79	4.17	0.02	.89	.00
	Pairs	5.00	5.52			
Number of different justification types (range 0-7)	Individuals	2.57	2.56	1.23	.28	.04
	Pairs	2.11	2.00			
Overall critical evaluation score (range 0-5)	Individuals	2.07	1.94	0.14	.71	.01
	Pairs	2.21	1.69			
Topic						
Number of relevant justifications* (range 0-20)	GMO	4.14	3.02	0.64	.43	.02
	Social Media	4.58	5.41			
Number of different justification types (range 0-7)	GMO	2.29	2.06	0.06	.81	.00
	Social Media	2.31	2.31			
Overall critical evaluation score (range 0-5)	GMO	2.57	1.51	1.02	.32	.04
	Social Media	2.04	1.84			

* For the analysis, variable was normalized, but original *Ms* and *SDs* are reported here.