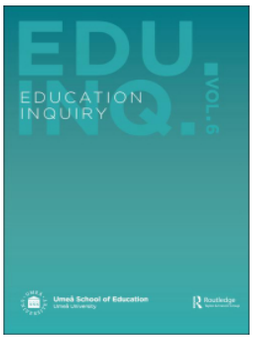

Education Inquiry

ISSN: 2000-4508 (Online) Journal homepage: www.tandfonline.com/journals/zedu20

# Boosting fact-checking in the classroom: verifying war photos and the pitfalls of overconfidence in education against disinformation

Thomas Nygren, Markus Al-Afifi & Carl-Anton Werner Axelsson

**To cite this article:** Thomas Nygren, Markus Al-Afifi & Carl-Anton Werner Axelsson (19 Sep 2025): Boosting fact-checking in the classroom: verifying war photos and the pitfalls of overconfidence in education against disinformation, Education Inquiry, DOI: 10.1080/20004508.2025.2558407

**To link to this article:** https://doi.org/10.1080/20004508.2025.2558407

© 2025 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group.

View supplementary material

Published online: 19 Sep 2025.

Submit your article to this journal

Article views: 347

View related articles

View Crossmark data

Full Terms & Conditions of access and use can be found at https://www.tandfonline.com/action/journalInformation?journalCode=zedu20

EDUCATION INQUIRY
https://doi.org/10.1080/20004508.2025.2558407

Routledge
Taylor & Francis Group

OPEN ACCESS Check for updates

# Boosting fact-checking in the classroom: verifying war photos and the pitfalls of overconfidence in education against disinformation

Thomas Nygren [a], Markus Al-Afifi [a] and Carl-Anton Werner Axelsson [a,b]

[a]Department of Education, Uppsala University, Uppsala, Sweden; [b]School of Innovation, Design and Engineering, Mälardalen University, Uppsala, Sweden

**ABSTRACT**

This study investigates the learning outcomes of a classroom-based intervention designed to help upper-secondary students verify the time and location of war-related photographs, thereby improving their ability to navigate visual disinformation, specifically concerning the war in Ukraine. The intervention engaged 192 Swedish students (ages 15–19) in three parts: (1) a teacher-led introduction featuring a video on war-related disinformation, (2) student activities with video-based modelling of lateral reading and reverse image search, and (3) a follow-up discussion to review correct answers. Using a pre-test-post-test design with both closed and open-ended survey items, the study assessed changes in students' fact-checking abilities, use of digital tools, and confidence in their judgements. Results show that students improved in identifying both in-context and out-of-context images and increased their use of reverse image search tools. However, the study also identified an important side effect: while high-performing students became more metacognitively aware, some low-performing students showed increased confidence despite unchanged or low performance, a pattern consistent with the Dunning-Kruger effect. These findings underscore the importance of designing educational interventions against disinformation that address both cognitive skill development and metacognitive calibration, particularly for students at risk of developing unwarranted confidence in their evaluative abilities.

**KEYWORDS**

Media literacy; misinformation; fact-checking; educational intervention; overconfidence

## 1. Introduction

Various scholars underscore misleading online information as a threat to democracy (e.g Ecker et al., 2024)., Similarly, the World Economic Forum (2024) has identified misinformation and disinformation[1] as the most significant short-term threats to contemporary societies across the globe. Visual disinformation is particularly potent due to its emotional appeal and perceived credibility, reinforcing the notion that "seeing is believing" (Iyer et al., 2014). While deepfakes produced with generative AI have gained attention, research and practitioners consistently report that simpler manipulations, such as mislabelled or decontextualised photos, are more widespread and impactful in wartime communication (Suarez, 2023; Toler, 2019). Consequently, in times of war, the spread of misleading information may intensify, which can have dire consequences for swaying public opinion and diluting public discourse with false information.

The importance of learning how to verify photos enters a more complex terrain when the problem of wartime mis- and disinformation can also be exaggerated with AI. In an era of generative AI, it is possible to generate and spread (dis)information rapidly, but the AI models have shown limited capacity in verifying photos (Chandrasekar & Jaźwińska, 2025). Today, all citizens need to be able to verify misleading images as part of a democratic self-defence – to not be manipulated and be able to make well-founded judgements about hot topics (Caulfield & Wineburg, 2023; Nygren, 2019).

The significance of teaching students how to identify dis- and misinformation similar to professional fact-checkers is today stressed by scholars (e.g. Caulfield & Wineburg, 2023; Ecker et al., 2022; Guess et al., 2020; Roozenbeek et al., 2023, 2024) and international organisations like the EU and UNESCO (Carlsson,

**CONTACT** Thomas Nygren thomas.nygren@edu.uu.se Department of Education, Uppsala University, Box 2136, 750 02 Uppsala, Sweden
Supplemental data for this article can be accessed online at https://doi.org/10.1080/20004508.2025.2558407.

© 2025 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group.
This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial License (http://creativecommons.org/licenses/by-nc/4.0/), which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited. The terms on which this article has been published allow the posting of the Accepted Manuscript in a repository by the author(s) or with their consent.

2019; European Commission, 2022a). More specifically, Wineburg et al. (2022) assert that students would benefit from mastering skills like "reverse image searches" p. 906). To counter manipulation, professional fact-checkers frequently employ open-source tools and reverse image search engines like Google, TinEye, Yandex, and Bing to determine the origin and temporal context of images (Aprin et al., 2022; Nygren et al., 2021). Previous research has highlighted how conducting reverse image searches can be part of successful educational interventions to support teenagers' fact-checking abilities (Axelsson et al., 2021) and older adults (Moore & Hancock, 2022). Conversely, interventions teaching reverse image strategies may also have little or no effect on people's evaluation of misinformation when using infographics to inform participants in online web panels of how to conduct reverse image searches (Qian et al., 2023, p. 12). Nor may interventions displaying reverse image searches against political misinformation in home settings affect participants' "ability to identify misinformation" (Badrinathan, 2021, p. 1337). Therefore, teaching students to utilise reverse image search like fact-checkers may be possible, but also challenging.

Adding to the challenge of ineffective educational interventions is the challenge of potential side effects from education against disinformation. Scholars identify that educational interventions against disinformation may make people overly critical of real news (Guess et al., 2020; Modirrousta-Galian & Higham, 2023) and make students cynical rather than critical and constructive (Haider & Sundin, 2022).

In response to the surge of misleading imagery during the early stages of the Russian invasion of Ukraine, we developed an educational digital resource informed by previous classroom-based media literacy research (Axelsson & Nygren, 2024; Axelsson et al., 2021). The digital resource included videos modelling expert practices, exercises encouraging lateral reading, and structured classroom discussions. This resource was designed in line with the principles of cognitive apprenticeship, offering students opportunities to observe, practice, and reflect on expert fact-checking strategies in an authentic classroom setting. By March 2022, more than 6,000 students had engaged with our digital resources, demonstrating a strong demand for practical and accessible tools to navigate wartime information. The present study builds on this outreach, systematically evaluating the intervention's impact on students' performance and metacognitive calibration.

Against this backdrop, the present research focuses on learning reverse image searching to evaluate visual misinformation related to armed conflicts. To better understand the potential and challenges of education against visual misinformation in relation to the contemporary context, our analysis centres on images from the ongoing war in Ukraine. We see that teaching students to evaluate misleading images from war is important, but also a challenge, where it is necessary to use a design that supports a cognitive apprenticeship (Axelsson & Nygren, 2024) and pay attention to unwanted side effects (Tay et al., 2023).

### 1.1. Framing of the study

This study aims to explore the potential and limitations of teaching secondary school students to evaluate visual disinformation using reverse image search, a method commonly employed by professional fact-checkers (Suarez, 2023). The educational intervention was designed within a cognitive apprenticeship framework, providing modelling, guided practice, and reflection to support students' fact-checking skills in an authentic learning environment (Collins et al., 1989). Drawing on recent concerns in the literature about both the efficacy and unintended consequences of disinformation education, this study also investigates a critical metacognitive dimension: the alignment – or misalignment – between students' actual performance and their confidence in evaluating war-related images. Specifically, the study addresses two key challenges. First, it examines whether students can improve their ability to identify contextual variables (e.g. time and location) of manipulated or decontextualised images through a structured classroom intervention. Second, it assesses whether such interventions might unintentionally lead to increased overconfidence. In doing so, the study contributes to research on effective skill development while also attending to potential psychological side effects, including unwarranted scepticism and inflated confidence.

By focusing on real-world examples from the war in Ukraine, a conflict marked by intense visual disinformation campaigns, this study provides insights into how fact-checking education can be both empowering and problematic, depending on how it is designed and for whom it is implemented. The decision to use this conflict as a focal point stems from its documented prominence in visual disinformation campaigns targeting European audiences (Wenzel et al., 2024). Although geographically distant, the Ukraine war represents a pedagogically relevant and empirically rich context through which students

encounter real-world examples of manipulated images. The spread of misleading videos and images has been a persistent problem in both past and present wars, where, for example, photos of suffering victims, war crimes, or battlefield victories are taken from unrelated contexts to support false claims (e.g. Häggström, 2022; Lytvynenko & Silverman, 2020). This use of decontextualised visuals to deliberately mislead continues in recent wars, including those in Ukraine and Gaza (Häggström, 2022; Khaldarova & Pantti, 2020; Seitz & Klepper, 2022; Suarez, 2023). These often include *cheap fakes*, such as out-of-context photos repurposed to provoke emotional responses or spread false claims (Fazio, 2020; Suarez, 2023).

## 2. Theoretical framework

### 2.1. Digital civic literacy

In this study, we build on the concept of *media and information literacy* (MIL), broadly defined as the knowledge and skills required to access, evaluate, analyse, and create information in digital environments (Aufderheide, 1993; Eshet, 2004; Hobbs, 2010). Following recent scholarship, we use MIL as an umbrella term encompassing multiple forms of digital literacy, including photo-visual literacy, lateral reading, and information literacy – competencies increasingly considered essential for meaningful civic engagement in a digitised democracy (Bateman & Jackson, 2024; European Commission, 2022b; Nygren & Guath, 2022). Critically evaluating digital content is often considered a foundational "survival skill" for 21st-century citizens.

From this perspective, MIL is not merely a technical skillset but a core component of *digital civic literacy*, which refers to the ability to responsibly and effectively navigate digital public spheres. This includes recognising and resisting disinformation, engaging in democracy with diverse perspectives, and verifying sources before sharing content. As such, teaching MIL is also an investment in civic agency, critical thinking, and democratic resilience (Nygren & Guath, 2022; Roozenbeek et al., 2023).

### 2.2. Cognitive apprenticeship in disinformation education

In designing educational interventions for digital civic literacy, this study draws on the theory of *cognitive apprenticeship* (Collins et al., 1989; Dennen, 2004). Cognitive apprenticeship emphasises the importance of making expert thinking visible and gradually transferring responsibility for complex tasks to learners.

In the context of digital fact-checking, this approach means helping students reason and act like professional fact-checkers. This includes demonstrating expert techniques such as reverse image search, guiding students through practice in authentic tasks, and supporting their reflection on both process and outcome. The cognitive apprenticeship model has shown promise in digital civic literacy interventions that enable learners to adopt expert-like strategies such as *lateral reading*, which involves verifying claims by consulting multiple independent sources rather than staying within the original site (e.g. Axelsson & Nygren, 2024; Wineburg et al., 2022).

While the full cognitive apprenticeship model includes elements of modelling, scaffolding, coaching, and reflection, this study concentrated primarily on two of these dimensions: modelling and reflection. The intervention featured video demonstrations where an expert fact-checker performed and verbalised the steps involved in using reverse image search tools to evaluate war-related photographs. This approach made tacit knowledge, specifically using reverse image search, explicit and observable to students. In this way, expert reasoning was modelled to students in a manner consistent with the principles of cognitive apprenticeship.

After completing their tasks, students were encouraged to reflect on their own verification process and learning, linking their experiences to the modelled strategies. However, unlike more scaffolded or teacher-led approaches, the intervention provided minimal real-time coaching or interactive support during the task phase. Instead, students engaged in self-regulated practice, applying the modelled strategies on their own in a classroom setting.

This selective implementation of cognitive apprenticeship, focused on exposure to expert strategies and post-task reflection, was designed to be scalable and feasible within regular classroom constraints. At the same time, this design choice may help explain the observed increase in overconfidence among some

students, particularly those who struggled with accuracy yet reported higher confidence post-intervention, as discussed further in the paper.

### 2.3. Metacognition and overconfidence

A central concern in this study is the role of *metacognition*—students' awareness and regulation of their own cognitive processes – in shaping how they engage with disinformation. Self-regulated learning involves setting goals, monitoring progress, and evaluating outcomes – processes generally associated with improved academic performance. However, when learners lack accurate self-assessment skills, self-regulated learning can inadvertently reinforce overconfidence (Bjork et al., 2013). While fact-checking strategies can be taught, students may not accurately assess their own competence, leading to either underconfidence (which may discourage engagement) or *overconfidence* (which may reinforce poor judgement). The latter is of particular concern in disinformation education, where students may believe they are effective fact-checkers without actually demonstrating the necessary skills (Nygren & Guath, 2019).

This phenomenon is related to the well-known *Dunning-Kruger effect*, which describes how individuals with lower ability tend to overestimate their competence (Kruger & Dunning, 1999). Such overconfidence can undermine efforts to promote responsible information practices and may be common among people with low levels of information literacy (Mahmood, 2016). Addressing these metacognitive challenges ensures that educational interventions against disinformation do not unintentionally foster a false sense of mastery among students with limited fact-checking abilities.

## 3. Previous research

### 3.1. The challenge of digital civic literacy among teenagers

Although MIL has received growing attention in policy and research, young people often struggle to apply these skills in real-world contexts. Large-scale assessments have shown that many students find it difficult to distinguish between factual and opinion-based claims or verify digital content's reliability (Breakstone, Smith, Wineburg, et al., 2021; Ku et al., 2019; OECD, 2021).

Furthermore, the challenge is not only technical, but also civic. A lack of digital civic literacy can weaken democratic resilience by leaving citizens more susceptible to manipulation and disinformation. Despite growing awareness of this issue, schools struggle to integrate effective instruction against disinformation into regular classroom practice, particularly outside controlled research settings (Roozenbeek et al., 2024).

### 3.2. Successful interventions based on cognitive apprenticeship

Despite these challenges, a growing body of research points to promising educational strategies to improve students' ability to evaluate online content. Interventions grounded in cognitive apprenticeship have been shown to help students adopt expert-like habits of verification, particularly through modelling and practising lateral reading strategies (Axelsson & Nygren, 2024; McGrew, 2020; Wineburg et al., 2022). Interventions that allow students to "see and do" fact-checking in ways that mirror real-world practice tend to yield better outcomes than abstract or information-based approaches (Nygren & Axelsson, in press).

Several studies have demonstrated the effectiveness of these interventions across age groups – from middle school (Artmann et al., 2023; Nygren & Axelsson, in press) to high school (Axelsson et al., 2021; Wineburg et al., 2022), university students (Breakstone, Smith, Connors, et al., 2021), and older adults (Moore & Hancock, 2022). Reverse image search, in particular, has been identified as a practical and impactful skill that can be taught across these contexts (Nygren et al., 2021; Qian et al., 2023).

Yet not all interventions succeed. Studies relying on infographics or one-off home-based activities have shown limited impact, especially on participants' ability to apply fact-checking strategies in politically charged or emotionally salient contexts (Badrinathan, 2021; Qian et al., 2023).

### 3.3. The problem of overconfidence in disinformation education

In the Swedish context, Nygren and Guath (2019) have found that high school students who expressed strong confidence in their digital source criticism often performed poorly on objective tests. This mirrors findings in other studies pointing to a general pattern of miscalibration, where low-performing individuals rate their abilities highly (Mahmood, 2016). These findings suggest that educational interventions should not only focus on skill acquisition but also support students in developing accurate self-assessment and intellectual humility (Prike et al., 2024; Tay et al., 2023).

## 4. The present study

Building on the theoretical perspectives of digital civic literacy, cognitive apprenticeship, and metacognition, this study investigates a classroom intervention aimed at supporting students' abilities to verify the authenticity of war-related visual content. We focus specifically on *reverse image search* to evaluate how students engage with visual disinformation related to the war in Ukraine.

The study focuses on two interrelated components of digital civic literacy: Students' ability to correctly identify the time and place of authentic versus out-of-context war images; and the alignment between students' confidence in their answers and their actual performance. We recognise that such alignment is crucial for supporting *intellectual humility*, an important disposition in democratic digital engagement (Nygren & Guath, 2019; Prike et al., 2024). At the same time, overconfidence, especially among low-performing students, may undermine critical information practices, contributing to poor judgement and false certainty (Bjork et al., 2013; Mahmood, 2016).

This leads to our guiding research questions: RQ1: To what extent does a short educational intervention improve students' ability to identify the time and place of war-related and out-of-context images using reverse image search? RQ2: How does the intervention affect students' confidence in their responses to visual disinformation? RQ3: What is the relationship between students' confidence and actual performance, particularly among high- and low-performing students?

We hypothesised that the intervention would enhance students' reverse image search skills and lead to greater alignment between confidence and competence. However, we also explored the possibility that improvements in self-reported confidence may diverge from actual ability, especially among confident students.

## 5. Methods

The intervention design in this study was grounded in insights from prior classroom-based interventions addressing misinformation (Axelsson et al., 2021). It was tailored to align with both the Swedish national curriculum, which mandates that students develop the ability to critically assess digital information (Swedish National Agency for Education, 2011), and with international guidelines from the EU and UNESCO that emphasise media and information literacy as a civic competency (European Commission, 2022a, 2022b; UNESCO, 2011). A central component of the intervention was a video modelling expert use of reverse image search, developed in line with research on digital cognitive apprenticeship and the benefits of demonstrating expert practices to support students' lateral reading skills (Axelsson & Nygren, 2024; Axelsson et al., 2021).

Given the documented prevalence of misleading images in online war reporting, the intervention included explicit instructions on conducting reverse image searches using freely available digital tools. Importantly, the intervention was implemented in authentic classrooms to test its effectiveness in real-world educational settings. This approach responds to the call for more ecologically valid research that acknowledges the complexity of classroom environments (Anderson & Shattuck, 2012; Brown, 1992; Shavelson et al., 2003) and aligns with recent appeals for rigorous classroom-based studies on educational interventions targeting misinformation (Roozenbeek et al., 2024).

### 5.1. Context and participants

In Sweden, media and information literacy have been part of compulsory schooling since 2011 (Sundin, 2015). The civics syllabus, which these participants study, underscores how all students should be able to

evaluate information from different sources and media formats (Swedish National Agency for Education, 2011). The Swedish Schools Inspectorate (2018, 2024) has criticised schools for not teaching students to scrutinise visual online information critically.

We recruited 234 students aged 15–19 across five different schools in three cities located in mid-Sweden (see Appendix F for more details about the participants). Due to time constraints, student absenteeism, and challenges in managing pre-test/post-test codes, we have completed and matched pre- and post-test scores from 192 students. The pre- and post-test assessments were conducted as part of the intervention, focusing on immediate outcomes rather than long-term effects. The teaching occurred in regular classroom settings, ensuring the intervention was integrated into the standard educational environment. All participating students had internet access to computers, providing a consistent digital infrastructure to support the intervention.

### 5.2. Lesson design

The lesson plan for teaching included: (1) a teacher-led introduction about the war in Ukraine and disinformation with a short video about the war, (2) student activities with a short introduction (Figure 1) and hands-on exercises (Figures 2 and 3) where students investigated a mix of authentic and misleading news with images from Ukraine and watched our modelling video on reverse image search, and (3) a teacher-led discussion about the tasks, including reviewing the correct answers (material available online in Swedish[2] and English[3]).

### 5.3. Intervention design and measures

We employed a pre- and post-test design, combining basic knowledge questions, photo identification, and self-reported measures (Appendix D). Inspired by pragmatism, we used a mixed-methods approach to conduct an exploratory study to better understand the potential effects of a new kind of intervention (Creswell & Creswell, 2022; Johnson & Onwuegbuzie, 2004). This study design allows for exploring the intervention's impact on student performance. The instructional intervention was designed to support


Figure 1. Introduction for students. Screenshot from the English version of the website.

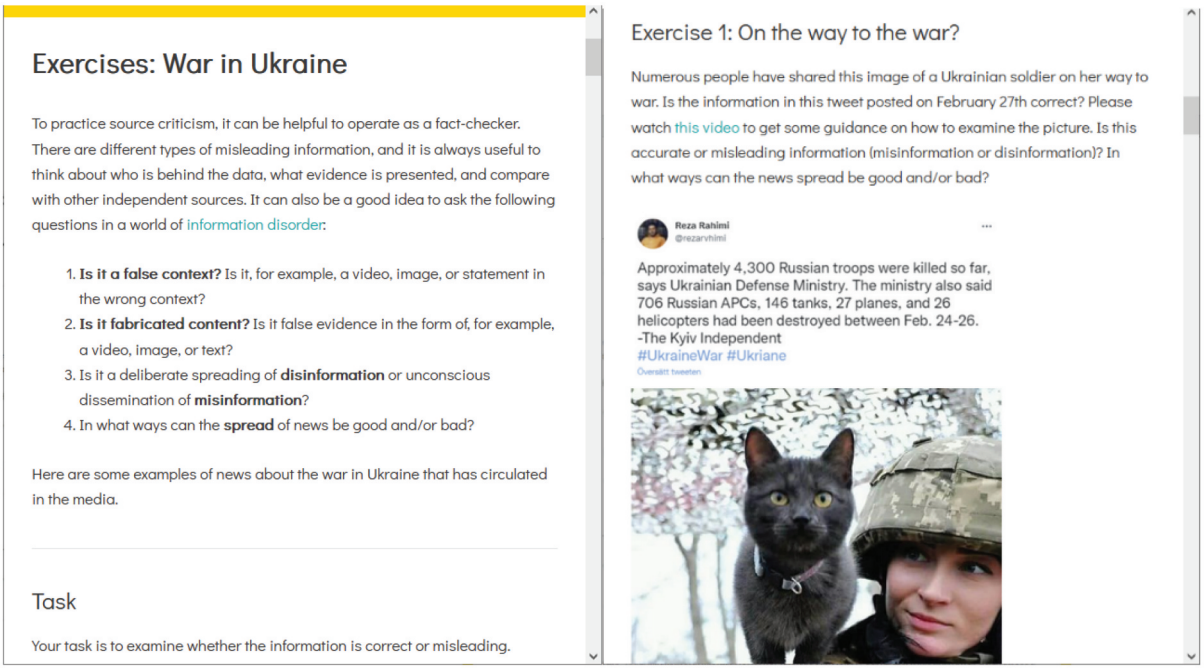


**Figure 2.** Exercises with instructions to students with a link to a video with an expert showing how to conduct a reverse image search. Screenshot from the English version of the website.

students' knowledge about the war, ability to fact-check images, and confidence. The intervention was limited to one and a half hours.

The pre-test and post-tests surveyed participants' background information, as well as measured their reverse image search ability, self-rated confidence in their response, and attitudes towards information credibility (Table 1). In the pre-test, we collected background information about media habits in relation to the war and trust in different types of information. In the pre-test and post-test, we asked the participants when and where four images of war were taken (similar but different in the pre-test and post-test), their confidence in their responses, use of digital aids such as reverse image search tools, and if they perceived that they had learned something new from the educational intervention (the latter concerns post-test only).

#### *5.3.1. Reverse image search to fact-check visual (mis)information*

As our main measure of reverse image search ability, students were asked to identify the dates and locations of four images before and after the intervention. Hence, we assessed students' ability to verify information of visual content by scoring their responses based on accurately identifying when the image was first posted online (date) and in what context (location) – in line with how fact-checkers do at times of war (Suarez, 2023). The scoring rubric was designed to reflect core fact-checking practices and to serve as a proxy for students' effective use of digital verification tools, specifically reverse image search. Since identifying specific dates and locations without such tools is generally difficult, successfully verifying the date and location of images out of context was interpreted as an indicator of digital tool use and a skill to identify disinformation. To assess students' ability to verify the date and location of war-related visual content, we developed a test consisting of four matched image pairs (1a/1b, 2a/2b, 3a/3b, and 4a/4b), each depicting thematically similar content across different contexts. The image pairs were selected to reflect types of visuals commonly found in online disinformation and included: hospitals in Mariupol, murals by Banksy, destroyed buildings in Gaza, and destroyed buildings in Kyiv (Appendix D). Two image pairs represented authentic, in-context content (Mariupol and Kyiv), while the other two (Banksy and Gaza) were out-of-context examples with images from other conflicts. All images were reviewed for thematic relevance, visual comparability, and technical quality to ensure difficulty and cognitive demand equivalence.

To control for order effects and balance exposure, students were randomly assigned in clusters to one of two test sequences. Circa half of the students ($N = 91$) in the a-groups received im ages 1a–4a in the pre-test

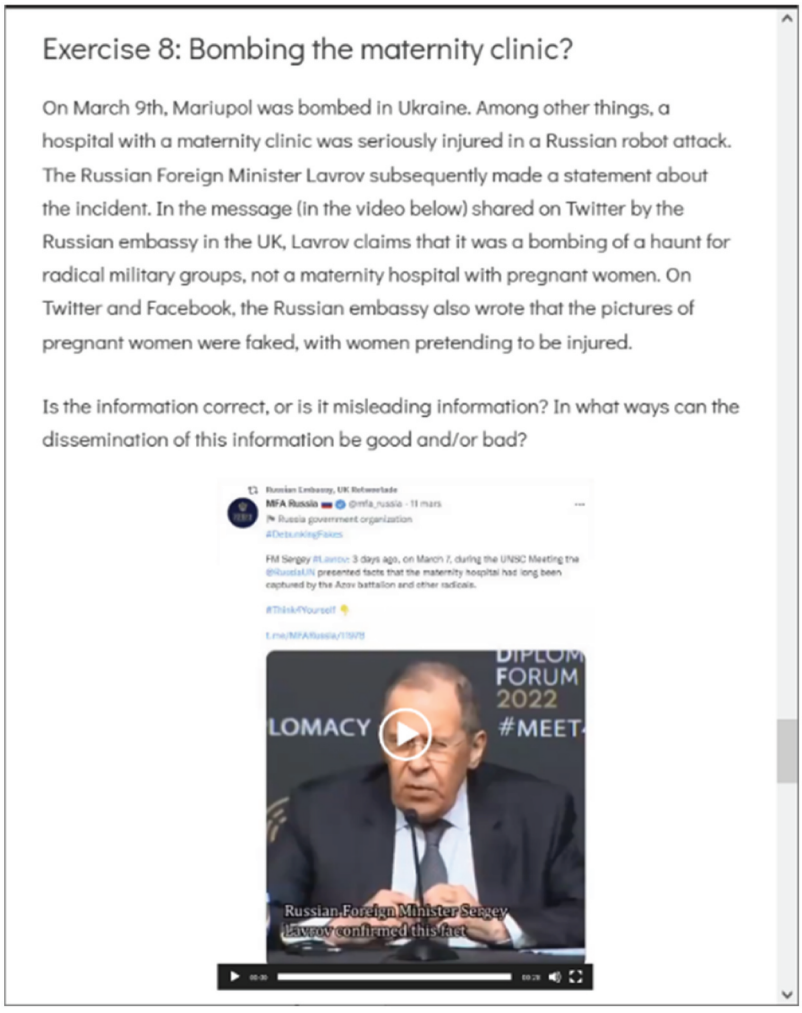


**Figure 3.** Instructions and exercises for students with a video to identify disinformation.

**Table 1.** Overview of the questionnaire in pre-test and post-test.

| Measured variable | Question category (number of items in pre/post) |
|---|---|
| ***Background variables*** | Gender (1/0) |
| | Grade Level (1/0) |
| | Upper-secondary school program (1/0) |
| | News consumption and media habits (4/0) |
| ***Self-reported categories*** | Yes/No (1/1) |
| Use of digital tools | Fact-checking certainty (5/5) |
| Confidence in own response (after test items of cognitive ability) | Yes/No (1/1) |
| Perceived learning (0/1) | |
| ***Cognitive ability*** | Date and location of image (4/4) |
| Reverse image search to fact-check visual information | |

and 1b–4b in the post-test; the other half in the b-groups ($N = 101$) completed the reversed sequence. This counterbalanced design enabled us to ensure fair comparisons. The image items were presented in a randomised order for each student to further reduce ordering bias. Importantly, all images were used solely as test items and were not included in any instructional materials, allowing us to isolate students' ability to apply fact-checking strategies without risk of recall or prior exposure. The design aimed to simulate authentic disinformation scenarios while assessing students' development in reverse image search and contextual reasoning under controlled classroom conditions.

To assess whether the image-based test items were comparably difficult across student groups, we analysed pre-test scores for students from the two clusters (a and b). Results showed similar performance at baseline, with a mean pre-test score of 2.25 for the a-groups and 2.01 for the b-groups (getting the images

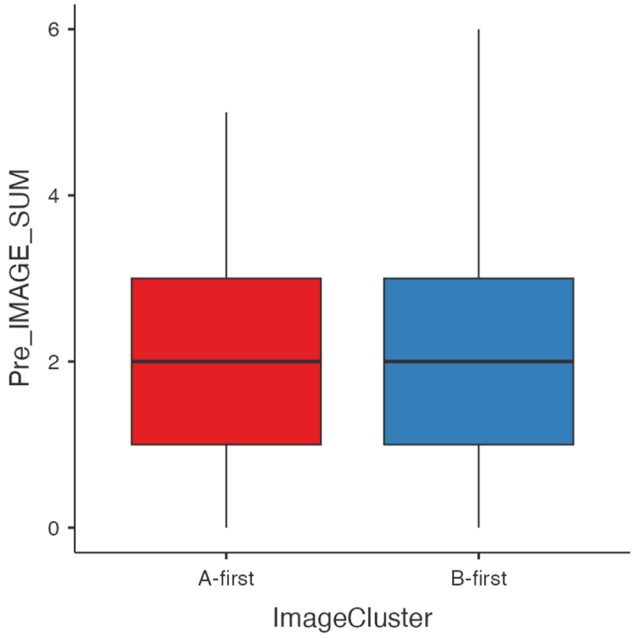


Figure 4. Image item difficulty based on pretest scores.

in reversed pre-test-post-test order), suggesting no substantial difference in initial task difficulty across cohorts (Figure 4).

For images from Ukraine, where familiarity due to ongoing media exposure could make identifying the country easier even without verification, we included an additional scoring dimension: correct identification of the specific city (e.g. Mariupol or Kyiv). These items allowed for a maximum of 3 points (date, country, city), while out-of-context images (e.g. Banksy murals and images from Gaza) allowed for a maximum of 2 points (date, country). We computed a composite score (0–10) of their ability to reverse image search by summing the scores of correctly identified date and location (pre-test $\omega = 0.595$, post-test $\omega = 0.796$).

The scoring system was pragmatically developed for this exploratory study, drawing on procedures used by professional fact-checkers at organisations such as Bellingcat, AFP, and B.B.C. Verify. By rewarding full points only for precise identification (particularly city-level accuracy in "in context" items), the rubric aimed to reduce the influence of guessing and increase the validity of inferences about actual digital tool use and reasoning processes. While the approach has limitations, it was designed to approximate real-world verification tasks and provide an objective yet practically feasible measure of students' digital civic literacy in visual disinformation.

### 5.3.2. Confidence in their fact-checking

Following each question about the identification of date and location in the test images, students were asked to self-assess their confidence in their response. This was measured using a 0–10 Likert-type scale, where 0 indicated "not at all confident" and 10 indicated "extremely confident". We computed a composite score (0–40) of their confidence to reverse image search by summing the scores of their self-rated certainty in their stated response to identifying the date and location of images (pre-test $\omega = 0.861$, post-test $\omega = 0.821$). This measure was used to assess students' metacognitive calibration, that is, the alignment between their perceived certainty and actual performance. Including confidence ratings for each item allowed us to explore patterns of overconfidence or intellectual humility concerning task difficulty and accuracy.

## 5.4. Statistical analysis

Because performance scores on the reverse image items and its composite scale did not follow a normal distribution (Shapiro-Wilk, $p < .001$), we used the non-parametric Wilcoxon signed-rank test (Field, 2017) to

investigate the interventional impact on reverse image search ability. Similarly, the confidence composite scale did not follow a normal distribution (Shapiro-Wilk, $p < .001$), warranting a non-parametric test (Appendix A).

Multiple linear regression was used to investigate the factors contributing to students' performance and confidence following the intervention. This statistical approach moves beyond simply establishing a statistical difference between pre- and post-test and allows for an in-depth examination of which predictors can be attributed to the potential change observed in the Wilcoxon signed rank tests. Variables were chosen based on the theoretical relevance of the key mechanisms at play in the intervention. In the first regression model, the post-test performance of reverse image search was set as the outcome variable, with post-test confidence, using digital tools as predictors and controlling for relevant covariates, such as baseline performance. In the second model, post-test confidence was regressed on post-test performance and use of digital tools, including covariates such as baseline confidence. By switching the dependent and independent variables in models 1 and 2, we can probe the relative effects of digital tools on post-test performance and post-test confidence, respectively, while also controlling for performance and confidence as well as covariates. In a third model, we regressed students' overconfidence, computed as the standardised difference between post-test performance and post-test confidence, on their use of digital tools in the post-test, baseline performance and confidence levels, performance group, and covariates. We created a factor based on the performance group, splitting participants by the median of the post-test performance score (Low, $n = 106$; High, $n = 86$), setting high-performers as the reference group. Accordingly, model 3 targets calibration by regressing an overconfidence index – defined as the difference between post-test confidence and post-test performance – on digital-tool use and performance group, with the same baseline controls as tested in models 1 and 2. These regression models (see Appendix C) were used to assess the relationship between performance scores and confidence levels, particularly among high- and low-performing students, while also controlling for relevant explanatory variables. Thus, this model setup tests isolates miscalibration of confidence relative to performance from different complementary angles. For details about the respective test values, see Appendices A–C.

## 6. Results

### 6.1. Impact on students' abilities to fact-check visual (mis)information from war

A majority of the participants correctly identified the country of origin when asked to locate photos from the war in Ukraine in the pre-test. Still, we saw a significant increase in locating the city when determining the origin of images from Mariupol post-intervention (Figure 5).

A Wilcoxon signed rank test revealed that the intervention significantly impacted students' performance in identifying photos from Mariupol ($p < .001$, $r = .68$; Appendix A). In contrast, the impact on students' ability to identify photos from Kyiv was non-significant ($p = .11$). While out-of-context images were generally more challenging for students to identify, the intervention significantly affected students' ability to detect the origin of these images. Improvements were statistically significant for both the images of Banksy murals ($p < .001$, $r = .93$) and destroyed buildings in Gaza ($p < .001$, $r = .85$). Still, only a third of the students correctly identified when and where the out-of-context photos were taken after the intervention. Overall, we found that the intervention positively impacted students' total scores on the four test items with a large effect size ($p < .001$, $r = .75$).

#### 6.1.1. The use of digital tools

Only a few students used digital tools to identify when and where the photos were taken pre-intervention, but this number tripled post-intervention ($n_{pre} = 34$, 18%; $n_{post} = 110$, 57%). To assess the statistical difference between this descriptive change, a Wilcoxon signed rank test was conducted to compare students use of digital tools before and after the intervention. The difference was confirmed to be statistically significant, $p = < .01$ with a strong effect size of $r = 0.86$.

### 6.2. Confidence change and overconfidence among low performers

The boxplot illustrates that confidence level change from the pre-test to the post-test (Figure 6). As reflected in the boxplot's median differences and distribution, a Wilcoxon signed rank test revealed that the

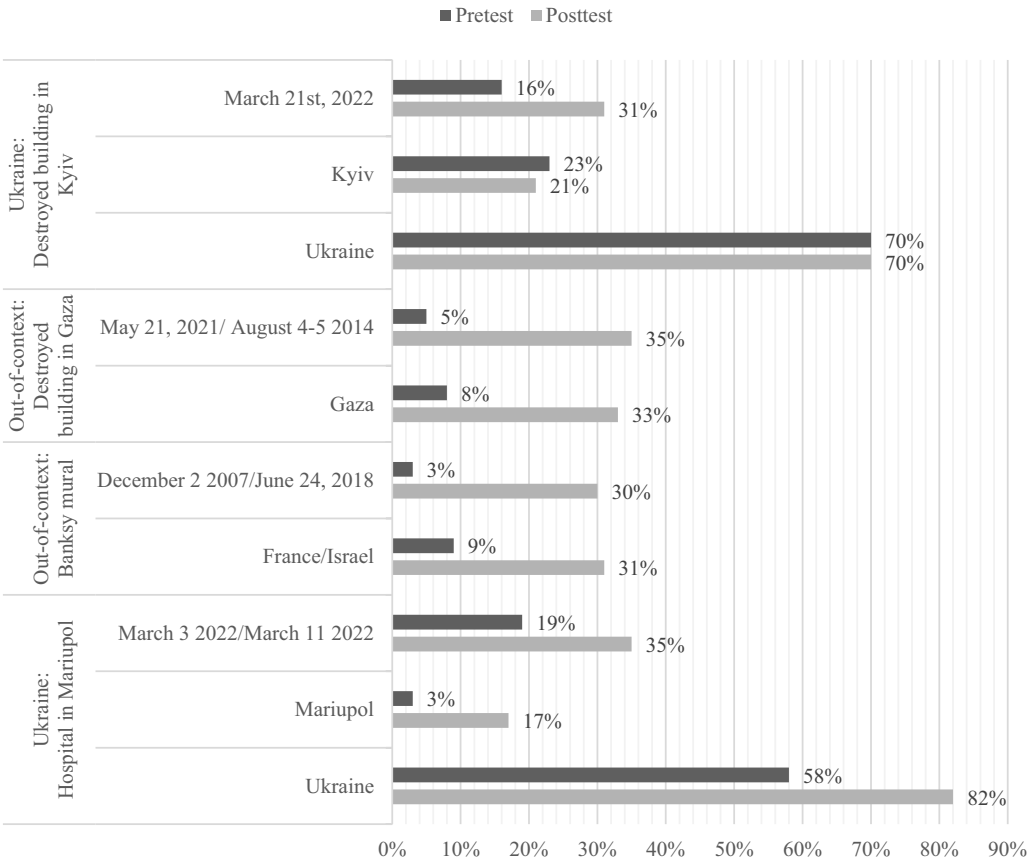


**Figure 5.** Students' identification of photos from the war in Ukraine and out-of-context images ($n$=192).

intervention significantly impacted students' confidence in their abilities to identify the date and location of the prompted images ($p < .001$, $r = .70$; Appendix A). Thus, the intervention significantly and positively affected students' confidence in their responses to visual disinformation.

To futher investage this confidence change, we probed the relationship between students self-rated confidence and their ability to identify the date and location of images. While confidence increased for many students post-intervention, these gains were not always aligned with improved performance. Based on participant scores in the pre-test regarding identifying the origins of images from the war (scored 0–3) and out-of-context images (scored 0–2), we found that the impact of the intervention varied between groups of low- and high-performing students. 108 students did not improve their accuracy between the pre- and post-test. These students still reported increased confidence on at least one post-test item, despite performing no better – or worse – than before (see Table 2). This pattern of overconfidence among low performers suggests that some students developed a false sense of mastery after engaging in the intervention.

In comparison, top-performing students, scoring the max of 2 out of 2 or 3 out of 3, showed decreased certainty across three out of four test items, hinting at increased intellectual humility (see Figure 7 and Appendix B, Table B2-B4). Most students in the middle groups grew more confident in their answers (Appendix B).

Figure 7 illustrates how students' confidence changed across image sets and performance levels. A clear pattern emerged: students with the lowest pre-test scores (Score 0) consistently reported increased certainty after the intervention, even when their post-test performance remained low or only slightly improved. In contrast, students with high pre-test scores (e.g. Score 2 or 3) often reported stable or even reduced certainty, particularly on the out-of-context (Mariupol and Banksy) items. This divergence highlights a possible overconfidence effect among low performers and increased metacognitive calibration among top performers. Certainty increased most among low performers, even in cases where performance remained low, indicating overconfidence. In contrast, some high-performing students showed decreased certainty post-intervention, possibly reflecting greater metacognitive awareness or intellectual humility.

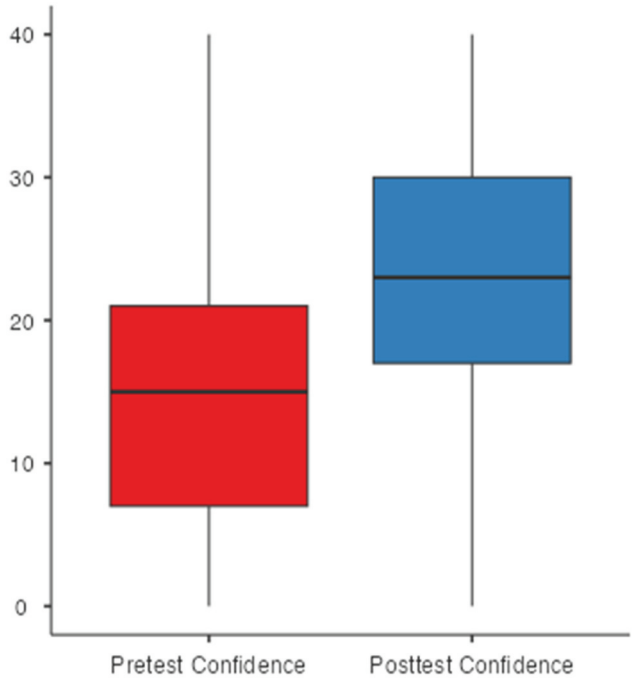


Figure 6. Box plot of students' confidence in pretest and posttest.

Table 2. Students with increased confidence in relation to incorrect responses.

| Confidence Changes | N | % |
|---|---|---|
| 0.00 The student did not increase his/her confidence in any incorrect answer | 86 | 44.8% |
| 1.00 The student increased his/her confidence in one question, even though the answer was incorrect | 78 | 40.6% |
| 2.00 The student increased his/her confidence on two questions, even though the answers were incorrect | 26 | 13.5% |
| 3.00 The student increased his/her confidence on three questions, even though the answers were incorrect | 1 | 0.5% |
| 4.00 The student increased his/her confidence on four questions, even though the answers were incorrect | 1 | 0.5% |

### 6.2.1. Reasoning among overconfident low-performers

To better understand this problematic relationship, we qualitatively analysed the responses from low-performing students who became more confident without improving their performance. Following our exploratory mixed-method approach, we further triangulated our findings, indicating a boost of over-confidence through a thematic analysis of responses from students who consistently scored low on a specific test item (0p to 0p) and those whose scores declined (e.g. 1p to 0p) but reported increased confidence in their answers.

Inspired by Braun and Clarke (2006), we identified three themes in the evolution of low-performing students' responses. (a) From vague to elaborate: In the pre-test, students provided vague responses, such as "in a hospital" or "last week". In the post-test, they transitioned to more detailed answers, though still incorrect, which may indicate overconfidence. (b) From hesitant to assertive: Some students moved from uncertain, hedged responses (e.g. "I think this picture . . . ") to direct, confident and incorrect statements, which could reflect overconfidence. (c) From guessing to acknowledging limitations: This group replaced guesses with admissions of not knowing, reflecting a deeper understanding of the complexity involved in identifying the origins of an image. This change likely does not indicate overconfidence. The adverse effect indicates that our intervention led to overconfidence among students with poor results.

## 6.3. Perceived, but not always actual learning

81% of the students reported learning something new from the educational intervention. They wrote that they, for instance, learned how "easy it is to be fooled", "how images and videos are easily used for the wrong purpose to support a claim", and "how to check a picture or video to see if it is accurate or not", and how to conduct "image searches" and use digital tools like Tineye. Students also stated that they learned to identify

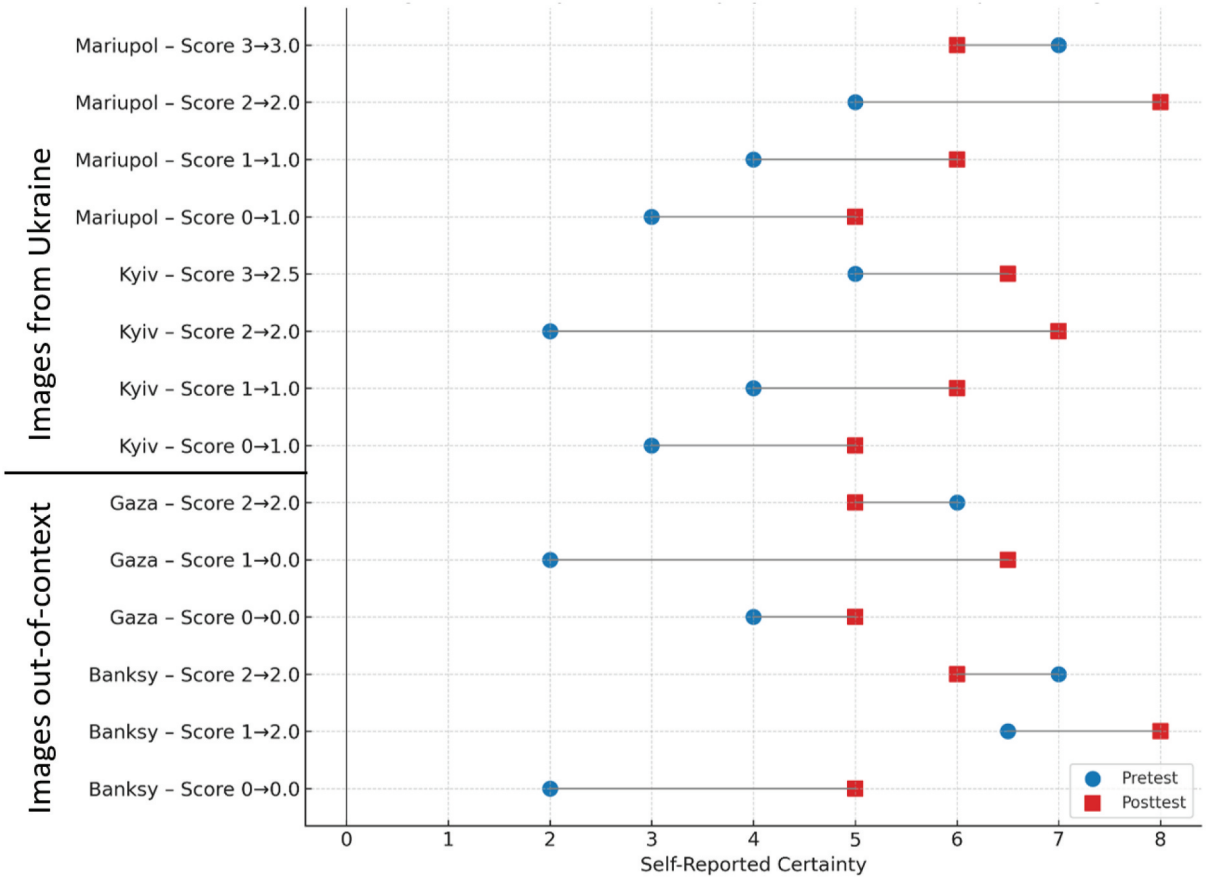


**Figure 7.** Changes in students' self-reported certainty by pretest performance group and image set. Note: Each line represents a subgroup of students categorized by pretest performance (e.g. Score 0→1). Blue circles indicate pretest certainty levels; red squares indicate posttest levels.

different types of misleading information. Although many perceived learning, 65% could not identify out-of-place images post-intervention. Further statistical analysis reveals that perceived learning had no statistically significant effect on post-test performance scores, $p = .21$, suggesting that the participants' perception of learning and actual learning were not aligned. This misalignment between perceived and actual learning contributes to understanding the observed overconfidence effect.

### 6.4. Predictors of individual differences in performance, confidence, and overconfidence

To identify which variables contributed to the change identified in the Wilcoxon tests, multiple regressions were conducted to examine the relationship between students' abilities to identify photos from the war and out-of-context images (performance scores) and confidence levels, particularly among high- and low-performing students, to gain deeper insight into the observed changes in performance and confidence (Table 3).

In model 1, we explored whether differences in post-test performance scores were associated with post-test confidence levels and digital tool use, controlling for baseline performance score, gender, and perceived learning. The results showed that confidence was a significant predictor of performance, $B = 0.04$, $p = .004$, indicating that, on average, students with higher self-perception of certainty in their performance tended to achieve better outcomes, suggesting alignment in students' metacognitive awareness. The regression results from model 1 also showed that using digital tools had no statistically significant effect on the post-test score of image corroboration, $p = > .05$. Hence, although the intervention significantly affected the use of digital tools for image corroboration, it did not affect the accuracy of their responses, on average. Among the control variables, baseline performance was the only significant predictor of post-test performance, $B =$

**Table 3.** Regressions of specified outcome variables on determinants.

| Variable | Model 1: Performance | Model 2: Confidence | Model 3: Overconfidence |
|---|---|---|---|
| Constant | 3.63** | 14.20** | −0.61 |
| | (0.50) | (2.61) | (0.24) |
| Performance | – | 1.06* | – |
| | | (0.36) | |
| Confidence | 0.04* | – | – |
| | (0.01) | | |
| Digital tool use | 0.33 | 6.09** | 0.49* |
| | (0.30) | (1.42) | (0.16) |
| Perceived learning | 0.26 | −1.01 | −0,16 |
| | (0.26) | (1.29) | (0.15) |
| Baseline Performance | 0.36** | −0.04 | −0.09* |
| | (0.06) | (0.34) | (0.04) |
| Baseline Confidence | 0.01 | 0.23** | 0.02* |
| | (0.01) | (0.06) | (0.01) |
| Performance Group | – | – | 0.74** |
| Low – High | | | (0.16) |
| Gender | 0.11 | −0.33 | −0.06 |
| Male – Female | (0.23) | (1.17) | (0.14) |
| Adj. $R^2$ | 0.78 | 0.51 | 0.16 |
| *N* | 190 | 190 | 190 |

Note. *$p < 0.05$; **$p < 0.01$. The table reports unstandardised regression coefficients with standard errors in parentheses.

0.36, $p = < .001$, indicating that students with better skills to identify images entering the intervention also maintained higher performance after the intervention, showing a cumulative learning effect.

In model 2, we tested whether post-test performance scores and digital tool use predicted post-test confidence levels, controlling for gender, perceived learning, and baseline confidence (Table 3). Post-test performance score was a significant predictor, $B = 1.06$, $p = .004$. This result shows that students who performed better also expressed more certainty in their answers, aligning with the results from model 1 about students' metacognitive awareness. In addition, digital tools had a statistically significant and positive effect on the post-test score of image corroboration, $B = 6.09$, $p = < .001$, suggesting that students who use digital tools felt more confident in their ability to identify images, even though the results from the model 1 revealed that using digital tools had no statistically significant on their actual performance. Among the control variables, baseline confidence was the only significant predictor of post-test confidence, $B = 0.23$, $p = < .001$, showing that students with better self-perception of their abilities to identify images entering the intervention also maintained higher confidence in their abilities after the intervention.

In model 3, we evaluated how performance group and digital tool use predicted students' overconfidence (Table 3). The performance group was a significant predictor, $B = 0.74$, $p = < .001$, showing that low-performers reported higher confidence than the high-performers after the intervention while controlling for pre-test confidence. Thus, relative to their actual skill level, low performers systematically reported inflated confidence in their image corroboration accuracy compared to high performers, indicating over-confidence. Furthermore, using digital tools was positively associated with increased overconfidence, $B = 0.49$, $p = .003$. Aligning with the direct effect of digital tool use on confidence and no significant effect on performance observed in the prior regression models, these results suggest that tool use made students feel more accurate, although their performance did not improve accordingly. Regarding the control variables, baseline confidence was positively associated with post-test overconfidence, $B = 0.02$, $p = .003$, indicating that existing misalignment in metacognitive awareness prior to the intervention carried throughout the intervention. Conversely, baseline performance was negatively associated with post-test overconfidence, $B = 0.09$, $p = .01$, suggesting that students who performed better on the pre-test were less overconfident in the post-test. This result shows that low performers, based on pre-test scores, were more overconfident in the post-test. Hence, baseline performance and confidence predicting post-test overconfidence suggest that early signs of low performance and low metacognitive awareness of image corroboration skills correlate with poorer metacognitive accuracy even after learning how to reverse image search.

In sum, the series of regression models demonstrates that perceived and actual performance were positively associated at the group level but that these signs of metacognitive awareness were conditional upon not splitting the sample by performance group. Model 3 contributed with nuanced insights into the relationship between performance and confidence, showing a disproportionate overconfidence

pattern among low performers when factoring in performance level. Additionally, the use of digital tools was not significantly associated with increased performance (model 1) but instead positively associated with confidence (model 2) and overconfidence (model 3). These results highlight the potentially problematic aspects of introducing new digital tools without making sure students understand how to use them correctly, which, as shown in this study, contributes disproportionately to confidence relative to performance. Together, our findings show that overconfidence peaks among the lower-performing students and that high-performing students have more accurate metacognitive awareness.

## 7. Discussion

This study set out to explore not only whether students could learn to verify the date and location of war-related images using reverse image search, but also how well their confidence in doing so aligned with their actual performance. This dual focus bridges cognitive skill development and metacognitive self-awareness in digital civic literacy. In a contemporary high-salience war context where dissemination of mis- and disinformation constitutes a significant threat to contemporary societies (World Economic Forum, 2024), this study has shown how a low-resource intervention can improve students' photo-verification skills. However, the regression results (Appendix C) also indicate a calibration risk. While post-test confidence predicted performance, digital-tool use predicted confidence more than corroboration accuracy, and a pattern of overconfidence was observed among lower performers. Our regression results thus highlight a complex interplay between the relative contributing role of digital tool use in improving performance and calibrating confidence. Methodologically, the study contributes a paired in/out-of-context item design, a fact-checking intervention grounded in professional practice, and a useful overconfidence index that teachers can use to track confidence alongside accuracy, clarifying when tool-supported instruction helps and where additional metacognitive scaffolding is needed. Our design, informed by cognitive apprenticeship and classroom experiences, enabled us to observe both cognitive and metacognitive outcomes in relation to a complex and socially significant task.

In line with RQ1, we find that a brief intervention rooted in cognitive apprenticeship can significantly enhance students' ability to verify the origin and context of visual content – a core component of digital civic literacy. Improvements were observed across both in-context and out-of-context images, with large effect sizes, particularly for less familiar or deceptive content. We also observe changes in confidence levels (RQ2) among students concerning their abilities to fact-check visual images from war. However, while most students reported greater confidence post-intervention, this growth was not uniform across groups. Our findings show that for many low-performing students, confidence increased even when performance did not, which we interpret as a pattern suggesting overconfidence (RQ3). The mismatch between students' actual and perceived ability, especially among those who started with low performance, highlights a critical metacognitive blind spot. In contrast, high-performing students often showed increased intellectual humility.

### *7.1. Learning to verify images with digital tools*

The wars in Ukraine and, more recently Gaza, highlight the urgent need to protect citizens from both misinformation and disinformation. In this context, our study shows that an educational intervention, including instructional materials, modelling videos, and fact-checking assignments using digital tools, can improve students' abilities to verify visual content through reverse image search. The combination of instructions, modelling videos, and hands-on assignments appears to have provided students with a structured framework for evaluating the credibility of images. This aligns with research emphasising how cognitive apprenticeship may support students' ability to fact-check (e.g. Axelsson & Nygren, 2024).

Our findings contrast those of Qian et al. (2023) and Badrinathan (2021), who reported limited or no effects of reverse image search interventions on misinformation evaluation in other contexts. Qian et al. (2023) conducted online panel studies where the interventions relied primarily on infographics as the method of instruction. While infographics can raise awareness, they lack the modelling and interactive elements necessary for skill development, which were essential in the present study. Badrinathan (2021)

examined interventions conducted in home settings, where challenges such as ensuring comprehension, engagement, and motivation likely diminished their effectiveness.

The educational intervention had the most significant impact on out-of-context test items, which included out-of-context photos. This finding highlights the usefulness of lateral reading with reverse image search, especially when one's prior knowledge is limited. However, the benefits of lateral reading are more restricted when the information to evaluate contains well-known facts, such as identifying a familiar image or the start date of the war in Ukraine. However, lateral reading and reverse image search can still improve confidence in correct judgements.

We find that digital tools and trustworthy information can form a constructive combination when navigating misinformation. One needs to be able to corroborate news with something trustworthy to make sound judgements.

Combining skills with subject-specific knowledge and informed trust seems constructive. Noting how students rarely use digital tools to evaluate information, even in a classroom setting where they are supposed to think critically, underscores the importance of incorporating these practices in classrooms and consistently applying them to foster a habit among students. Knowing what sources to trust and following credible news outlets may be a central aspect for navigating (mis)information (Altay et al., 2023), and the importance of news in classrooms needs to be further investigated in future research. Correct and updated knowledge, in combination with an informed trust in sources, may also be key to not going down a problematic rabbit hole when using lateral reading on new and hot topics. What we see in this study is that proper knowledge and source trust may be key, especially since only a few students seem to have a spontaneous habit of conducting lateral reading (Axelsson & Nygren, 2024; Axelsson et al., 2021).

### 7.2. Confidence gains, potential miscalibration, and problematic side-effects

Additionally, the intervention seems to have boosted overconfidence among students who were poor at fact-checking. We define overconfidence here as a metacognitive mismatch: when students report increased confidence without corresponding performance gains. This pattern, which affected more than half of the low-performing students, echoes findings in cognitive psychology on the Dunning-Kruger effect (Kruger & Dunning, 1999). It also aligns with research on self-regulated learning, which warns that surface-level familiarity can create an illusion of competence (Bjork et al., 2013). In this context, the intervention may have given students some experience evaluating images without sufficient depth to achieve accuracy, leading to misplaced confidence. While this study was informed by the cognitive apprenticeship framework (Collins et al., 1989; Dennen, 2004), it employed only selected elements, primarily expert modelling through instructional videos and reflection following task completion. This limited implementation likely contributed to uneven outcomes. In particular, the absence of scaffolding and real-time coaching meant that students received little guided support as they attempted to apply complex verification strategies independently. This may have left lower-performing students vulnerable to misapplication of strategies and metacognitive miscalibration, as evidenced by the rise in overconfidence among some subgroups. These findings underscore a broader challenge in education against disinformation: without structured opportunities for feedback and dialogic support, modelling alone may be insufficient to foster both accurate performance and calibrated self-assessment. Hence, educational interventions against misinformation may induce problematic side effects requiring consideration in research and practice. Without clear benchmarks or corrective feedback, students may confuse task completion with task mastery.

In contrast, students who obtained higher scores became less confident about their knowledge and skills, indicating increased intellectual humility. This may be interpreted as a positive impact in light of research linking intellectual humility to people's abilities to evaluate (mis)information (Prike et al., 2024). This outcome may reflect an adaptive calibration process, in which students gain not only knowledge but also insight into the limits of their knowledge.

Our findings emphasise the need to thoroughly evaluate interventions to support students' media and information literacy to avoid adverse side effects. Previous research has noted potential challenges of unwanted effects from interventions against misinformation (Guess et al., 2020; Haider & Sundin, 2022; Modirrousta-Galian & Higham, 2023). Here, we identify another crucial challenge for education against misinformation in practice: lessons on lateral reading techniques may boost overconfidence in

groups of students struggling to navigate out-of-context images. This side-effect must be noted since we know there is a problematic divide in people's abilities to navigate misinformation (Nygren & Guath, 2022), and overconfidence may be a real problem among some students (Mahmood, 2016; Nygren & Guath, 2019). However, the specific factors influencing students' overconfidence remain unanswered. We see some clues in the transition to more detailed or assertive responses among students who develop an inflated sense of competence. This aligns with research indicating that surface-level engagement with skills or concepts can create an illusion of competence (Bjork et al., 2013). The self-regulated nature of fact-checking tasks, such as identifying misleading images, may contribute to overconfidence. These tasks typically require students to make independent judgements, monitor their reasoning, and assess the credibility of visual content without immediate corrective feedback. In the absence of clear benchmarks or external guidance, students may rely on superficial cues, such as task familiarity or intuitive plausibility, rather than rigorous verification strategies. This can lead to an *illusion of mastery*, where learners feel confident because they completed the task, not because they did so accurately (Bjork et al., 2013).

### 7.3. Implications for teaching and learning

Our findings highlight the need for future research to refine interventions that not only improve technical skills like reverse image search but also cultivate critical self-awareness and reflective thinking in the face of misinformation. Interventions against misinformation should be empirically investigated, with a particular focus on students who are most vulnerable to false information. This may help us better understand this challenge and make targeted adjustments to lateral reading interventions to support underperforming students.

Guidelines and teachers must ensure that low-performing students are identified and that they learn to acknowledge and appreciate their misunderstandings. This is a complicated balancing act since we want students to become confident in their abilities but not overconfident. It is central to support students' metacognitive processes and not support illusions of learning (e.g. Bjork et al., 2013). Teachers must pay attention to and support struggling students, and interventions may need to be tailored to vulnerable groups.

The results suggest that effective instruction for digital civic literacy, particularly when addressing complex tasks such as reverse image verification, requires more than expert modelling. While videos can inspire and inform students, many learners may struggle to internalise and apply strategies without interactive scaffolding and formative feedback. To mitigate the risks of overconfidence and misapplication, instruction should ideally incorporate guided practice, peer collaboration, and real-time coaching, allowing students to articulate their reasoning and receive targeted support. Teachers may also consider using self-assessment tools or structured reflection prompts to help students evaluate not only the correctness of their judgements, but also the basis for their confidence. In this way, instruction can better cultivate both the technical skills and the metacognitive awareness needed to navigate visual disinformation in emotionally and politically charged contexts.

It is evident that simply asking students if they learned what they were supposed to learn cannot be used as evidence that they did learn more complex skills. Thus, we suggest that teachers include test questions in low and high-stakes testing that evaluate students' abilities to fact-check online information (using, for instance, out-of-context images) and combine them with ratings of how confident the students are in their responses. This will give the teacher insights into individual students' abilities and (over)confidence.

## 8. Limitations

One important limitation of the study is that McDonald's omega for the reverse image search scale in the pre-test fell below the conventional threshold for acceptable internal consistency. The suboptimal reliability suggests that the results should be interpreted cautiously and warrants a recommendation for future research to refine the scale further to strengthen the internal validity. However, while McDonald's omega was relatively low in the pre-test, it reached acceptable internal consistency levels in the post-test. Hence, the variability in internal consistency suggests that participants'

understanding of the target construct got clearer post-intervention, aligning well with our efforts to develop students' understanding of how to use reverse image search to verify information from war-based images.

As this quasi-experimental study lacked a control group, the identification of causal inference is limited. While the intervention was novel and brief, reducing the likelihood of maturation effects, changes in performance and confidence cannot be definitively attributed to the intervention. Still, the observed covariance offers a valuable first step towards establishing causal links.

Moreover, this study does not provide insight into how common the associated issues with over-confidence are from interventions against disinformation, nor does it offer any explanations regarding how effective this intervention is with other groups and types of misinformation, tools, or in the long term. We would need multiple studies across diverse settings to determine these effects better. How to best support and sustain intellectual humility is also a question for future research. The transfer of knowledge from school to the world outside the classroom remains a challenge, so future research should be conducted in naturalistic environments and try to link learning in schools to students' habits at home. Still, this study highlights the potential of implementing a new educational design with tools for fact-checking in a challenging situation and the importance of considering students' (over)confidence and paying special attention to vulnerable groups.

## Notes

1. In line with Wardle's & Derakhshan (2017) typology, we distinguish between misinformation, which refers to false or misleading content shared without intent to deceive; disinformation, which denotes deliberately false information spread to mislead; and malinformation, which involves genuine information used to harm, often by taking content out of context.
2. In Swedish. Information for teachers: https://nyhetsvarderaren.se/ukraina/larare/, introduction: https://nyhetsvarderaren.se/ukraina/, excercises: https://nyhetsvarderaren.se/ukraina/ovningar/.
3. In English. Information for teachers: https://nyhetsvarderaren.se/ukraine/teachers/, introduction: https://nyhetsvarderaren.se/ukraine/, https://nyhetsvarderaren.se/ukraine/exercises/.

## Acknowledgments

We are truly grateful for all the support from teachers and students participating in the study. We would also like to sincerely thank Ullrich K. H. Ecker for valuable comments.

## Disclosure statement

No potential conflict of interest was reported by the author(s).

## Funding

This research was funded by the Swedish Institute for Educational Research, Skolforskningsinstitutet, (2020-00009).

## Notes on contributors

*Thomas Nygren* is Professor in History and Civics Education at the Department of Education at Uppsala University. His research interests focus on history and civics education, the digital impact on education, critical thinking and human rights education.

*Markus Al-Afifi* is a PhD candidate in Civics Education at the Department of Education at Uppsala University. His research examines how social science teaching can foster students' intercultural competence as part of schools' democratic mandate, with a focus on educational practices that promote social cohesion and increase political and social tolerance toward immigrants in an increasingly pluralistic Swedish society.

*Carl-Anton Werner Axelsson* is an associate lecturer in didactics of digital literacy and lifelong learning in computer science at the School of Innovation, Design and Engineering at Mälardalen University. His doctoral degree is in human–computer interaction and his research is focused on computer-supported learning and decision making.

## ORCID

Thomas Nygren http://orcid.org/0000-0003-1884-3252
Markus Al-Afifi http://orcid.org/0009-0006-0941-8595
Carl-Anton Werner Axelsson http://orcid.org/0000-0002-5194-9430

## CRediT

Thomas Nygren: Conceptualisation, Methodology, Investigation, Formal analysis, Funding acquisition, Project administration, Resources, Supervision, Writing – original draft.
Markus Alafifi: Investigation, Formal analysis, Writing – original draft.
Carl-Anton Werner Axelsson CRediT: Software, Visualisation, Writing – review & editing

## Declaration of use of generative AI

During the preparation of this work, the first author used ChatGPT and Grammarly in order to improve the readability of some passages. After using ChatGPT, the first author reviewed and edited the content as needed and takes full responsibility for the content of the publication.

## Data availability statement

All materials needed to replicate this study are available on this site: https://osf.io/8euvc/and this site: https://nyhetsvarderaren.se/ukraina/

## Ethics

This research was approved by the Swedish Ethical Review Authority (no. 2023–02188–01). All participants were at least 15 years old. Informed about the purpose of the study, they agreed to participate in line with national and international ethical recommendations.

## References

Altay, S., Berriche, M., Heuer, H., Farkas, J., & Rathje, S. (2023). A survey of expert views on misinformation: Definitions, determinants, solutions, and future of the field. *Harvard Kennedy School Misinformation Review*, *4* (4), 1–34. https://doi.org/10.37016/mr-2020-119

Anderson, T., & Shattuck, J. (2012). Design-based research: A decade of progress in education research? *Educational Researcher*, *41*(1), 16–25. https://doi.org/10.3102/0013189x11428813

Aprin, F., Chounta, I.-A., & Hoppe, H. U. (2022). "See the image in different contexts": Using reverse image search to support the identification of fake news in Instagram-like Social media. In S. Crossley & E. Popescu (Eds.), *Intelligent tutoring systems* (pp. 264–275). Springer International Publishing. https://doi.org/10.1007/978-3-031-09680-8_25

Artmann, B., Scheibenzuber, C., & Nistor, N. (2023). Elementary school students' information literacy: Instructional design and evaluation of a pilot training focused on misinformation. *The Journal of Media Literacy Education*, *15*(2), 31–43. https://doi.org/10.23860/JMLE-2023-15-2-3

Aufderheide, P. (1993). *Media literacy. A report of the national leadership conference on media literacy*. ERIC.

Axelsson, C. A. W., Guath, M., & Nygren, T. (2021). Learning how to separate fake from real news: Scalable digital tutorials promoting students' civic online reasoning. *Future Internet*, *13*(3), 60. https://doi.org/10.3390/fi13030060

Axelsson, C. A. W., & Nygren, T. (2024). The advantage of videos over text to boost adolescents' lateral reading in a digital workshop. *Behaviour & information technology*, 1–15. https://doi.org/10.1080/0144929X.2024.2308046

Axelsson, C. A. W., Nygren, T., Roozenbeek, J., & van der Linden, S. (2021). *Bad news* in the civics classroom: How serious gameplay fosters teenagers' ability to discern misinformation techniques. *Journal of Research on Technology in Education*, *57*(5), 1–27. https://doi.org/10.1080/15391523.2024.2338451

Badrinathan, S. (2021). Educative interventions to combat misinformation: Evidence from a field experiment in India. *The American Political Science Review*, *115*(4), 1325–1341. https://doi.org/10.1017/S0003055421000459

Bateman, J., & Jackson, D. (2024). *Countering disinformation effectively: An evidence-based policy guide*. https://carnegieendowment.org/research/2024/01/countering-disinformation-effectively-an-evidence-based-policy-guide?lang=en

Bjork, R. A., Dunlosky, J., & Kornell, N. (2013). Self-regulated learning: Beliefs, techniques, and illusions. *The Annual Review of Psychology*, *64*(1), 417–444. https://doi.org/10.1146/annurev-psych-113011-143823

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology*, *3*(2), 77–101. https://doi.org/10.1191/1478088706qp063oa

Breakstone, J., Smith, M., Connors, P., Ortega, T., Kerr, D., & Wineburg, S. (2021). Lateral reading: College students learn to critically evaluate internet sources in an online course. *Harvard Kennedy School Misinformation Review*. https://doi.org/10.37016/mr-2020-56

Breakstone, J., Smith, M., Wineburg, S., Rapaport, A., Carle, J., Garland, M., & Saavedra, A. (2021). Students' civic online reasoning: A national portrait. *Educational Researcher*, *50*(8), 505–515. https://doi.org/10.3102/0013189X211017495

Brown, A. L. (1992). Design experiments: Theoretical and methodological challenges in creating complex interventions in classroom settings. *Journal of the Learning Sciences*, *2*(2), 141–178. https://doi.org/10.1207/s15327809jls0202_2

Carlsson, U. (Ed.). (2019). Understanding media and information literacy (MIL) in the digital age: A question of democracy. In *Department of journalism*. Media and Communication (JMG), University of Gothenburg, UNESCO.

Caulfield, M., & Wineburg, S. (2023). *Verified: How to think straight, get duped less, and make better decisions about what to believe online*. University of Chicago Press.

Chandrasekar, A., & Jaźwińska, K. (2025). Why AI models are bad at verifying photos. *Columbia journalism review*. Retrieved September 1, 2025, from https://www.cjr.org/tow_center/why-ai-models-are-bad-at-verifying-photos.php

Collins, A., Brown, J. S., & Newman, S. E. (1989). Cognitive apprenticeship: Teaching the crafts of reading, writing, and mathematics. In L. Resnick (Ed.). *Knowing, learning, and instruction: Essays in Honor of Robert Glaser* (pp. 453–494). Routledge.

Creswell, J. W., & Creswell, J. D. (2022). *Research design: Qualitative, quantitative, and mixed methods approaches*. SAGE Publications.

Dennen, V. P. (2004). Cognitive apprenticeship in educational practice: Research on scaffolding, modeling, mentoring, and coaching as instructional strategies. In D. H. Jonassen & M. Driscoll (Eds.), *Handbook of research on educational communications and technology* (pp. 813–828). Lawrence Erlbaum Associates.

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*(1), 13–29. https://doi.org/10.1038/s44159-021-00006-y

Ecker, U. K. H., Roozenbeek, J., van der Linden, S., Tay, L. Q., Cook, J., Oreskes, N., & Lewandowsky, S. (2024). Misinformation poses a bigger threat to democracy than you might think. *Nature*, *630*(8015), 29–32. https://doi.org/10.1038/d41586-024-01587-3

Eshet, Y. (2004). Digital literacy: A conceptual framework for survival skills in the digital era. *Journal of Educational Multimedia & Hypermedia*, *13*(1), 93–106.

European Commission. (2022a). *Final report of the Commission expert group on tackling disinformation and promoting digital literacy through education and training: Final report*. Directorate-General for Education, Y. S. & Culture. Publications Office of the European Union. https://doi.org/10.2766/283100

European Commission. (2022b). *Guidelines for teachers and educators on tackling disinformation and promoting digital literacy through education and training*. Directorate-General for Education, Y. S. & Culture. Publications Office of the European Union. https://doi.org/10.2766/28248

Fazio, L. (2020, February 14). *Out-of-context photos are a powerful low-tech form of misinformation*. The Conversation. http://theconversation.com/out-of-context-photos-are-a-powerful-low-tech-form-of-misinformation-129959

Field, A. (2017). *Discovering statistics using IBM SPSS statistics+ webassign*. Sage Publications.

Guess, A. M., Lerner, M., Lyons, B., Montgomery, J. M., Nyhan, B., Reifler, J., & Sircar, N. (2020). A digital media literacy intervention increases discernment between mainstream and false news in the United States and India. *Proceedings of the National Academy of Sciences*, *117*(27), 15536–15545. https://doi.org/10.1073/pnas.1920498117

Häggström, A. (2022, February 26). 6 fejkade bilder från invasionen i Ukraina. *Expressen*. https://www.expressen.se/nyheter/6-fejkade-bilder-fran-invasionen-i-ukraina/

Haider, J., & Sundin, O. (2022). *Paradoxes of media and information literacy: The crisis of information*. Routledge.

Hobbs, R. (2010). *Digital and media literacy: A plan of action. A white paper on the digital and media literacy recommendations of the Knight Commission on the information needs of communities in a democracy*. ERIC.

Iyer, A., Webster, J., Hornsey, M. J., & Vanman, E. J. (2014). Understanding the power of the picture: The effect of image content on emotional and political responses to terrorism. *Journal of Applied Social Psychology*, *44*(7), 511–521. https://doi.org/10.1111/jasp.12243

Johnson, R. B., & Onwuegbuzie, A. J. (2004). Mixed methods research: A research paradigm whose time has come. *Educational Researcher*, *33*(7), 14–26.

Khaldarova, I., & Pantti, M. (2020). Fake news: The narrative battle over the Ukrainian conflict. In Stuart Allan, Cynthia Carter, Stephen Cushion, Lina Dencik, Inaki Garcia-Blanco, Janet Harris, Richard Sambrook, Karin Wahl-Jorgensen, Andy Williams (Eds.). *The future of journalism: Risks, threats and opportunities* (pp. 228–238). Routledge.

Kruger, J., & Dunning, D. (1999). Unskilled and unaware of it: How difficulties in recognizing one's own incompetence lead to inflated self-assessments. *Journal of Personality & Social Psychology*, *77*(6), 1121–1134. https://doi.org/10.1037/0022-3514.77.6.1121

Ku, K. Y. L., Kong, Q., Song, Y., Deng, L., Kang, Y., & Hu, A. (2019). What predicts adolescents' critical thinking about real-life news? The roles of social media news consumption and news media literacy. *Thinking Skills and Creativity*, *33*, 100570. https://doi.org/10.1016/j.tsc.2019.05.004

Lytvynenko, J., & Silverman, C. (2020, January 8). *People are spreading false and unverified information about Iran's missile attack on US bases in Iraq. Don't Be fooled*. BuzzFeed News. https://www.buzzfeednews.com/article/janelytvynenko/false-and-unverified-information-iraq-attack

Mahmood, K. (2016). Do people overestimate their information literacy skills? A systematic review of empirical evidence on the Dunning-Kruger effect. *Communications in Information Literacy*, *10*(2), 3. https://doi.org/10.15760/comminfolit.2016.10.2.24

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

Modirrousta-Galian, A., & Higham, P. A. (2023). Gamified inoculation interventions do not improve discrimination between true and fake news: Reanalyzing existing research with receiver operating characteristic analysis. *Journal of Experimental Psychology: General*, *152*(9), 2411–2437. https://doi.org/10.1037/xge0001395

Moore, R. C., & Hancock, J. T. (2022). A digital media literacy intervention for older adults improves resilience to fake news. *Scientific Reports*, *12*(1), 6008. https://doi.org/10.1038/s41598-022-08437-0

Nygren, T. (2019). *Fakta, fejk och fiktion: Källkritik, ämnesdidaktik, digital kompetens. [Fact, fake and fiction: Source criticism, subject didactics, digital competence]*. Natur & Kultur.

Nygren, T., & Axelsson, C.-A. W. (in press). Don't tell me how to fact-check; show me, and let me try! A media literacy intervention with sixth-graders. *Journal of Media Literacy Education*. osf.io/agqy3_v2

Nygren, T., & Guath, M. (2019). Swedish teenagers' difficulties and abilities to determine digital news credibility. *Nordicom Review*, *40*(1), 23–42. https://doi.org/10.2478/nor-2019-0002

Nygren, T., & Guath, M. (2022). Students evaluating and corroborating digital news. *Scandinavian Journal of Educational Research*, *66*(4), 549–565. https://doi.org/10.1080/00313831.2021.1897876

Nygren, T., Guath, M., Axelsson, C. A. W., & Frau-Meigs, D. (2021). Combatting visual fake news with a professional fact-checking tool in education in France, Romania, Spain and Sweden. *Information*, *12*(5), 201. https://doi.org/10.3390/info12050201

OECD. (2021). *21st-century readers: Developing literacy skills in a digital world, PISA*. OECD Publishing. https://doi.org/10.1787/a83d84cb-en

Prike, T., Holloway, J., & Ecker, U. K. H. (2024). Intellectual humility is associated with greater misinformation discernment and metacognitive insight but not response bias. *Advances.in/Psychology*, *2*, e020433. https://doi.org/10.56296/aip00025

Qian, S., Shen, C., & Zhang, J. (2023). Fighting cheapfakes: Using a digital media literacy intervention to motivate reverse search of out-of-context visual misinformation. *Journal of Computer-Mediated Communication*, *28*(1), zmac024. https://doi.org/10.1093/jcmc/zmac024

Roozenbeek, J., Culloty, E., & Suiter, J. (2023). Countering misinformation: Evidence, knowledge gaps, and implications of current interventions. *European Psychologist*, *28*(3), 189–205. https://doi.org/10.1027/1016-9040/a000492

Roozenbeek, J., Remshard, M., & Kyrychenko, Y. (2024). Beyond the headlines: On the efficacy and effectiveness of misinformation interventions. *Advances.in/Psychology*, *2*(1), e24569. https://doi.org/10.56296/aip00019

Seitz, A., & Klepper, D. (2022, February 25). Propaganda, fake videos of Ukraine invasion bombard users. *AP news*. https://apnews.com/article/russia-ukraine-technology-europe-media-social-media-123c7975a879b89b85c06877f1f12908

Shavelson, R. J., Phillips, D. C., Towne, L., & Feuer, M. J. (2003). On the science of education design studies. *Educational Researcher*, *32*(1), 25–28. https://doi.org/10.3102/0013189X032001025

Suarez, E. (2023). *Bbc expert on debunking Israel-Hamas war visuals: "The volume of misinformation on Twitter was beyond anything I've ever seen*. Reuters. https://reutersinstitute.politics.ox.ac.uk/news/bbc-expert-debunking-israel-hamas-war-visuals-volume-misinformation-twitter-was-beyond

Sundin, O. (2015). Invisible search: Information literacy in the Swedish curriculum for compulsory schools. *Nordic Journal of Digital Literacy*, *10*(4), 193–209.

Swedish National Agency for Education. (2011). Swedish curriculum for upper secondary schools. https://www.skolverket.se/undervisning/gymnasieskolan/laroplan-program-och-amnen-i-gymnasieskolan/laroplan-gy11-for-gymnasieskolan

Swedish Schools Inspectorate. (2018). Undervisning om källkritiskt förhållningssätt i svenska och samhällskunskap, årskurs 7-9 [Teaching a critical approach to sources in swedish and civics, grades 7-9]. https://www.skolinspektionen.se/beslut-rapporter-statistik/publikationer/kvalitetsgranskning/2018/undervisning-om-kallkritiskt-forhallningssatt-i-svenska-och-samhallskunskap/

Swedish Schools Inspectorate. (2024). Undervisning om källkritiskt förhållningssätt. Teaching a critical approach to sources. https://www.skolinspektionen.se/beslut-rapporter/publikationer/kvalitetsgranskning/2024/kallkritik/

Tay, L. Q., Lewandowsky, S., Hurlstone, M. J., Kurz, T., & Ecker, U. K. (2023). A focus shift in the evaluation of misinformation interventions. *Harvard Kennedy School Misinformation Review*. https://doi.org/10.37016/mr-2020-124

Toler, A. (2019, December 26). *Guide to using reverse image search for investigations*. Bellingcat. https://www.bellingcat.com/resources/how-tos/2019/12/26/guide-to-using-reverse-image-search-for-investigations/

UNESCO. (2011). *Media and information literacy: Curriculum for teachers. United Nations Educational, Scientific and Cultural Organization*. UNESCO. http://unesdoc.unesco.Paris,France.org/images/0019/001929/192971e.pdf

Wardle, C., & Derakhshan, H. (2017). *Information disorder: Toward an interdisciplinary framework for research and policymaking*. Council of Europe.

Wenzel, M., Stasiuk-Krajewska, K., Macková, V., & Turková, K. (2024). The penetration of Russian disinformation related to the war in Ukraine: Evidence from Poland, the Czech Republic and Slovakia. *International Political Science Review*, *45*(2), 192–208. https://doi.org/10.1177/01925121231205259

Wineburg, S., Breakstone, J., McGrew, S., Smith, M. D., & Ortega, T. (2022). Lateral reading on the open internet: A district-wide field study in high school government classes. *Journal of Educational Psychology*, *114*(5), 893–909. https://doi.org/10.1037/edu0000740

World Economic Forum. (2024). *Global risks report 2024*. Retrieved January 15, 2024, from https://www.weforum.org/publications/global-risks-report-2024/