Computers in Human Behavior 161 (2024) 108385

Contents lists available at ScienceDirect

Computers in Human Behavior

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

Storifying instructional videos on online credibility evaluation: Examining engagement and learning

Riikka Anttonen [a], Kristian Kiili [a], Eija Räikkönen [b], Carita Kiili [a,*]

[a] Faculty of Education and Culture, Tampere University, P.O. Box 700, 33014, Tampere, Finland
[b] Faculty of Education and Psychology, University of Jyväskylä, P.O. Box 35, 40014, Jyväskylä, Finland

ARTICLE INFO

Handling editor: Nicolae Nistor

Keywords:
Online evaluation
Credibility evaluation
Instructional videos
Multimedia learning
Emotional design
Situational interest

ABSTRACT

This study used video-based instruction to promote adolescents' online credibility evaluation skills and to examine whether storification of instructional videos can maintain students' situational interest and positive activating epistemic emotions during a four-lesson intervention better than non-storified instructional videos. The study also investigated whether storified instruction can enhance students' learning of online credibility evaluation skills better than non-storified instruction. The learning content of the videos was equivalent, differing only in terms of storification. Students participated in either a non-storified ($n = 135$) or storified ($n = 115$) video-based instruction (4 x 45-min lesson). In each of the first three lessons, students watched an instructional video explaining one credibility evaluation strategy. In the fourth lesson, the video introduced a case requiring students to apply the taught strategies. After watching the video, students' situational interest and positive activating epistemic emotions (i.e., curiosity and excitement) were measured. Then, students practiced the taught strategies with learning tasks. Students' credibility evaluation skills were measured before and after the intervention with an online credibility evaluation task. The latent growth curve model showed that non-storified and storified videos similarly maintained students' situational interest. No decline was observed. However, a small decline in positive activating epistemic emotions was observed in both instructional conditions. Students' credibility evaluation skills improved equally across instructions. The results underline that professionally produced videos following several multimedia design principles can trigger and maintain situational interest and enhance students' online credibility evaluation skills without the need to incorporate additional storified elements.

1. Introduction

Evaluating the credibility of online information is a complex process. To determine whether online information is credible, readers must consider multiple aspects and their relations and how they affect text credibility (Barzilai et al., 2020; Forzani et al., 2022). Given these high demands, it is not that surprising that many adolescents struggle with online credibility evaluation (Breakstone et al., 2021; Forzani et al., 2022; Hämäläinen, 2023). In addition, adolescents may not always be that motivated to invest the cognitive effort that online credibility evaluation requires (Paul et al., 2017). This concern indicates that educators should develop instruction that considers both skill and will to ensure adolescents are adequately equipped to encounter the challenges of the current epistemic climate of online spaces (Chinn et al., 2021).

Although several intervention studies aimed at teaching adolescents credibility evaluation strategies have been conducted (e.g., Braasch et al., 2013; Bråten et al., 2019; Hämäläinen et al., 2023; Pérez et al., 2018), their instructional design mainly focused on cognitive aspects, with the exception of Barzilai et al. (2023). Barzilai et al. (2023) used game-based learning to engage and motivate students to learn evaluation strategies that would help them to identify both accurate and inaccurate information. As there is growing evidence that motivation and emotions are drivers of effective learning (Loderer et al., 2020), they should be considered to a greater extent when designing instruction intended to promote online credibility evaluation.

One promising way to increase students' affective and cognitive engagement with learning content is emotional design that incorporates engaging features (e.g., warm colors, rounded shapes, facial anthropomorphisms, narratives) into the instructional materials (Mahler & Mayer, 2024; Um et al., 2012). We consider storification to be one type

* Corresponding author.
E-mail addresses: riikka.anttonen@tuni.fi (R. Anttonen), kristian.kiili@tuni.fi (K. Kiili), eija.m.raikkonen@jyu.fi (E. Räikkönen), carita.kiili@tuni.fi (C. Kiili).

https://doi.org/10.1016/j.chb.2024.108385
Received 25 May 2024; Received in revised form 14 July 2024; Accepted 25 July 2024
Available online 2 August 2024

of emotional design that can potentially promote sense-making of learning materials due to increased interest and positive epistemic emotions. By storification we mean "contextualizing instruction, using fictitious characters as instructors, and incorporating playful elements to create a frame story that stimulates learner engagement" (Kiili et al., under review). The frame story can be created through metaphors or analogies that are regarded as a fantasy component of intrinsically motivating instruction (Malone, 1981). According to Malone (1981), a frame story or emotionally engaging fantasy can activate learners' existing knowledge, helping them to interpret the learning content.

In this study, we designed a four-lesson intervention to promote sixth graders' online credibility evaluation skills. To examine the impact of storified instruction, we created two sets of instructional materials: non-storified (control version) and storified (enhanced version) materials. The materials included instructional videos that taught credibility evaluation strategies and learning tasks with which students could practice the strategies and reflect on their learning. Non-storified videos did not include any specific context, characters, or playful elements. In contrast, storified videos used a detective metaphor to create a frame story by contextualizing online credibility evaluation as detective practices, using detective characters as instructors, and incorporating playfulness into the videos (see Fig. 1).

Specifically, we examined whether storified instructional videos could maintain students' situational interest and positive activating epistemic emotions (i.e., curiosity and excitement) during the online credibility evaluation intervention better than non-storified instructional videos. We also investigated whether video-based storified instruction enhances the learning of online credibility evaluation skills more than non-storified video-based instruction.

2. Theoretical framework

2.1. Online credibility evaluation

The internet is a vital information resource for adolescents in and out of school. In the school context, students can use the internet to construct knowledge (Wiley et al., 2009), explore different perspectives on controversial issues (Barzilai & Weinstock, 2020), and write from multiple online texts (Kirkpatrick & Klein, 2016). As on the internet, accurate and inaccurate information co-exist (Braasch & Graesser, 2019), evaluating credibility is crucial for effective learning from online information (Forzani et al., 2022; Leu et al., 2013).

Online credibility evaluation refers to readers' decision-making about what and whom to believe during reading and learning on the internet (Stadtler & Bromme, 2014). To determine the credibility of online information, readers can employ several evaluation strategies as outlined in the Bidirectional Model of First- and Second-Hand Evaluation Strategies (Barzilai et al., 2020). First-hand evaluation strategies can be used to evaluate the accuracy of content by comparing it to one's prior topic knowledge (Forzani, 2018; Lucassen et al., 2013), prior beliefs (van Strien et al., 2016), or to the content of other credible texts (Osborne & Pimentel, 2022). Further, readers can evaluate content by considering the quality of argumentation in texts. They can focus on the logic of evidence–claim relations, the persuasiveness of arguments, and whether evidence is based on scientific methodology (Iordanou et al., 2019). For example, readers can identify which types of evidence (e.g., research evidence, anecdotal evidence) authors offer and evaluate whether such evidence supports the presented claim (Kuutila et al., 2024; List, 2024).

When employing second-hand evaluation strategies, readers can focus on determining the trustworthiness of the source (i.e., author or publisher). Readers can identify authors' credentials and affiliations to make inferences about their expertise—that is, their knowledge and competence to provide accurate information on the topic at hand (Hendriks et al., 2016; Stadtler & Bromme, 2014). When evaluating the author's intentions, readers can consider whether the author offers the highest-quality information, without being driven by personal interests or agendas (Hendriks et al., 2015). By assessing author expertise and intentions, readers can develop a more comprehensive picture of the credibility of the texts than by focusing on the textual content alone (Barzilai et al., 2020; Forzani et al., 2022).

Importantly, the Bidirectional Model of First- and Second-Hand Evaluation Strategies emphasizes that judgments about the content and the source are reciprocal in nature (Barzilai et al., 2020). This means that evaluations of content affect judgments of source credibility, and evaluations of sources affect judgments of content. Unfortunately, many students tend to rely on superficial evaluation strategies without adequately considering the credibility of both the content and the source (Coiro et al., 2015; McGrew, 2021). Research also suggests that online credibility evaluation reflects the attributes of source texts (i.e., credibility and text genre), requiring two latent abilities: confirming the credibility of more credible texts and questioning the credibility of less credible texts, the latter being more challenging for younger readers (Kiili et al., 2023).

2.2. Instructional videos

As one form of multimedia learning material (Mayer et al., 2020), instructional videos are designed to help people learn specific content or skills (Fiorella & Mayer, 2018). Instructional videos can include on-screen text and/or instructor narration and visual elements, such as images, animations, and slides. When it comes to visuals, instructional videos may or may not feature an on-screen instructor (a human or a pedagogical agent).

2.2.1. Theoretical grounding of instructional videos

Empirical research has started to establish principles for designing effective educational videos (Beege et al., 2019; Fyfield et al., 2022; Mayer, 2021; Mayer et al., 2020).According to a recent literature review by Fyfield et al. (2022), the Cognitive Theory of Multimedia Learning

Fig. 1. Example screenshots of non-storified (Left-hand side) and storified (Right-hand side) videos explaining different types of evidence (personal experiences, expert knowledge, and research knowledge). *Note*. Actors: Iina Hautala and Aleksi Kauppinen.

(CTML) and related design principles (Mayer, 2014) constitute the most commonly used theoretical basis of instructional video design. The CTML assumes that learners process visual and auditory information in separate channels, each limited in capacity. Further, the CTML emphasizes that deep learning occurs when learners engage in active cognitive processing during learning. Therefore, instructional designers aim to optimize the use of learners' visual and auditory channels to facilitate active cognitive processing of learning content without overloading their information processing capacity.

When learning from instructional videos, learners' cognitive resources are allocated among essential, generative, and extraneous processing (Mayer, 2014, 2020). Essential processing is needed to hold and manipulate incoming information in working memory. The complexity of the learning content and learners' prior knowledge affects essential processing demands. Generative processing, which is facilitated by learners' motivation, is needed to make sense of incoming information. Finally, extraneous processing, caused by instructional design, refers to unnecessary cognitive processing that does not serve instructional goals. For example, decorative enrichment details included in instructional videos increase extraneous processing and may thus harm learning (Pink & Newton, 2020; Sundararajan & Adesope, 2020; Wang et al., 2020). The aim of instructional design is to optimize learning by managing essential processing, increasing generative processing, and minimizing extraneous processing.

Recently, scholars have extended the CTML to explain how motivational, affective, and social elements influence learning. Cognitive-affective models of learning (e.g., Mayer, 2020; Plass & Kaplan, 2016) posit that learners try harder to understand learning content when they experience positive emotions while learning. For example, the Integrated Cognitive-Affective Model of Learning with Multimedia (ICALM) outlines how cognitive processes are inseparably intertwined with affective processes, such as emotions or moods (Plass & Kaplan, 2016). The model highlights that learning material that can evoke learners' affect involving some appraisal can enhance their motivation and interest, followed by deeper processing and comprehension of new information.

Previous research suggests that emotional design features are worth integrating into learning materials. Techniques such as employing warm colors or incorporating anime or narrative storytelling into learning material have been shown to enhance learners' engagement and improve learning outcomes (Fyfield et al., 2022; Mahler & Mayer, 2024; Um et al., 2012). Anthropomorphism (i.e., incorporating human-like features into non-human elements of learning material) has also been found to have positive effects on learners' engagement and learning outcomes (Schneider et al., 2019; see meta-analyses by Brom et al., 2018; Wong & Adesope, 2021). However, the motivational effects of emotional design tend to be stronger for younger children (Brom et al., 2018). Regarding learners' cognitive resources, emotional design is a promising method of fostering learners' generative processing (Mahler & Mayer, 2024).

2.2.2. Instructional videos with on-screen instructors

In the present study, we focused on instructional videos with on-screen human instructors, which have become popular, particularly in higher education and industry (Henderson & Schroeder, 2021). A recent meta-analysis (Alemdag, 2022) found that the instructor's presence did not affect learning, but it did increase learners' motivation, enjoyment, and cognitive load. However, a systematic review by Polat (2023) reported mixed results concerning cognitive effects of the instructor's presence. Notably, although the instructor's presence can provide social cues that can enhance learners' motivation and cognitive engagement (Fiorella & Mayer, 2018), it can also increase unnecessary cognitive and affective processing demands that may hinder learning (Sweller et al., 2019).

Moreover, the instructor's appearance, such as dress style (Beege et al., 2019) and emotional expressions (Lawson et al., 2021a), may influence learners' cognitive, motivational, and affective processes. Regarding emotional expressions, Lawson et al. (2021a) found that when instructors displayed positive emotions, they were perceived as more engaging, more credible, and more likely to facilitate learning, when compared to instructors who exhibited negative emotions. Furthermore, a study by Schrader et al. (2021) indicated that the instructor's gender, and gender matching between the instructor and learner, had no impact on learning, cognitive load, or situational interest.

Findings from research on instructional videos with on-screen instructors align with cognitive-affective models of learning (e.g., Mayer, 2020; Plass & Kaplan, 2016), proposing that learners may exert greater effort to understand learning content when experiencing positive emotions. However, on-screen instructors and emotional design elements should not cause excessive extraneous cognitive load. In fact, the cost–benefit model of cognitive load (Skulmowski & Xu, 2022) suggests that the cost of extraneous processing of on-screen instructors and emotional elements should be lower than its motivational and generative processing benefits.

2.3. Situational interest and epistemic emotions

Motivation and emotions are essential drivers of learning. Motivation directs behavior and can also be experienced as interest (Elliot, 2023). In simple terms, emotions can be defined as affective episodes induced by a specific stimulus. Both cognition and emotions are involved in motivational processes, and their interaction may sustain motivation during learning (Schunk, 2023). According to Plass and Kaplan (2016), instructional design may evoke and support specific combinations of emotions and cognitions, which in turn serve as motivating forces that influence how learners interact with a learning environment or whether they disengage from learning activities. In the present study, we focused on situational interest and epistemic emotions as context-specific motivational and emotional factors that can vary in different situations (Pekrun & Marsh, 2022).

2.3.1. Situational interest

Interest is an essential motivational factor that has been shown to be a beneficial driver of learning (Elliot, 2023; Harackiewicz et al., 2000). In our study, we focused on situational interest for two reasons. First, situational interest can enhance cognitive and emotional engagement in learning (Hidi & Renninger, 2020). Second, it can be promoted via the instructional design (Renninger et al., 2019). Hidi and Renninger (2006, p. 113) defined situational interest as "focused attention and the affective reaction that is triggered in the moment by environmental stimuli, which may or may not last over time." Thus, situational interest can fluctuate during, for example, a specific learning task (Fulmer & Tulis, 2013; Koskinen et al., 2023).

Hidi and Renninger (2006) divided situational interest into triggered and maintained situational interest. *Triggered situational interest* can be sparked by environmental features and is typically externally supported (Hidi & Renninger, 2006). Specific captivating elements in learning materials and task instructions, such as enriched materials with warm, colorful elements (Brom et al., 2018; Um et al., 2012), positive, personalized language (Mayer et al., 2004), and anime stories (Mahler & Mayer, 2024), can trigger students' situational interest. Also, texts that are vivid or include narratives or surprising content with which readers can identify can trigger situational interest, making the texts more memorable (Hidi, 2001; Logtenberg et al., 2011). However, once situational interest has been triggered, it may or may not be maintained (Hidi & Renninger, 2006).

Maintained situational interest refers to a state of interest involving targeted and focused attention on a task for an extended period (Hidi & Renninger, 2006). Meaningful tasks and personally relevant activities can facilitate maintained situational interest (Harackiewicz et al., 2000). Maintained situational interest differs from triggered situational interest

mainly in terms of duration (Hidi & Renninger, 2006). Previous research has shown that instructional features may or may not maintain situational interest (Renninger et al., 2019). For example, in one study, interesting decorative illustrations triggered learners' situational interest but did not maintain it (Magner et al., 2014).

2.3.2. *Epistemic emotions*

According to Pekrun and Linnenbrink-Garcia (2012), emotions are pervasive in academic settings, and they can influence learners' engagement and performance. Previous research distinguished between different types of academic emotions, central among them being achievement emotions and epistemic emotions (Pekrun et al., 2018; Pekrun and Linnenbrink-Garcia, 2012a). According to Pekrun and Linnenbrink-Garcia (2012), epistemic emotions differ from achievement emotions in terms of their object. The object of achievement emotions relates to success or achievement in academic activities, whereas the object of epistemic emotions concerns the knowledge-generating qualities of academic activities. Regarding the present study, learners may experience epistemic emotions when processing the learning content of instructional videos, particularly if the provided information contradicts their previous conceptions and experiences (Pekrun et al., 2017, 2018).

Pekrun et al. (2017) have distinguished seven epistemic emotions: surprise, curiosity, enjoyment, confusion, anxiety, frustration, and boredom. These epistemic emotions can be categorized according to their valence (positive or negative) and strength of their physiological arousal (activating or deactivating). In this study, we focused on positive activating epistemic emotions (i.e., curiosity and enjoyment) because they tend to support learning (Park et al., 2015; Pekrun et al., 2017; Pekrun & Linnebrack-Garcia, 2014; Vilhunen et al., 2022). More precisely, epistemic curiosity can be defined as yearning for knowledge or the desire to acquire precise information that closes the learner's knowledge gap (Loewenstein, 1994; Shin & Kim, 2019). Common characteristics that trigger curiosity include uncertainty, complexity, and novelty (Hidi & Renninger, 2020). Regarding uncertainty, insufficient prior knowledge does not trigger curiosity, while excessive prior knowledge may lead to boredom. Learners can experience enjoyment

	Description		Video screenshots	
	Non-storified videos	Storified videos	Non-storified	Storified
Context				
Setting	No specific setting	Funky Facts detective office, which assists customers in determining which online texts can be trusted. Funky Facts has its own brand, which is depicted in its logo and thematic tune.		
Venue	No specific venue. Videos' background color is neutral beige.	Decorated detective office with detective tools and mysterious atmosphere.		
Characters				
Instructor	Instructor is restrained and professional.	Instructors are detectives with distinct personalities and principles for conducting their detective work, as depicted in their detective names.		

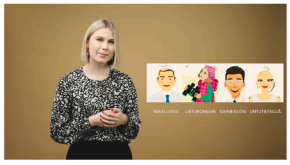
	Description		Video screenshots	
	Non-storified videos	Storified videos	Non-storified	Storified
Playfulness				
Communication and language	Communication is monologic and positive. Instructor uses neutral language.	Communication is dialogic and positive. Detectives use playful and humorous language. For example, online credibility evaluation strategies are framed as clues.		
Visuals	Instructor uses pictures and graphics to demonstrate the strategies. Plain text is used to signal the strategy.	Detectives use funny elements, such as finger puppets, to demonstrate strategies. Text on the mobile phone screen is used to signal the strategy.		
Sounds	Background music is neutral.	Background music is rhythmic. Certain sound effects are exaggerated to highlight, for example, suspicious intentions.		

Fig. 2. Characteristics of non-storified and storified instructional videos.

when curiosity is satisfied, i.e., uncertainty is reduced (Chevrier et al., 2019), when problems are solved, or when confusion is resolved (D'Mello & Graesser, 2012; Pekrun & Linnenbrink-Garcia, 2012). Furthermore, when experiencing enjoyment during a learning task, the learner's attention may be more task-directed (Pekrun et al., 2002).

3. The present study

For this study, we designed two video-based interventions (4 x 45-min lessons) aimed at promoting sixth graders' online credibility evaluation skills (evaluation of the author's expertise, the author's intentions, and the quality of evidence). Both interventions were implemented by classroom teachers. In this quasi-experimental study, we employed a value-added research design (e.g., Mayer, 2019) to examine the effects of emotional design in instructional videos. This study advances previous research by using storification as a unique emotional design approach.

First, we developed a basic version of the video-based instruction, including instructional videos and related learning tasks. In the videos, an instructor taught credibility evaluation strategies without any specific character or context. Second, we developed an extended version of the basic intervention by storifying it (see Fig. 1). In the storified video-based instruction, we used a detective metaphor to create a frame story. Students participated in the Funky Facts detective school (contextualizing the instruction by using a frame story), in which two detectives taught online credibility evaluation strategies through a dialogue (using fictitious characters as instructors) and with playful examples (incorporating playful elements to stimulate learner engagement) (see Fig. 2 for more details).

The rationale of the study stemmed from the theoretical contradiction that the emotional design of learning materials can have either beneficial or detrimental effects on learning. Some theoretical approaches argue that the use of unnecessary elements in learning materials, often referred to as seductive details, may distract learners' attention from the relevant learning content and undermine learning outcomes (Sweller, 1988; see meta-analysis by Sundararajan & Adesope, 2020). On the other hand, from an emotional design perspective (e.g., Mahler & Mayer, 2024; Wong & Adesope, 2021), the use of emotional design elements can foster motivation and engagement and thereby improve performance and learning outcomes. To contribute to this line of research with a focus on teaching online credibility evaluation, two specific research questions (RQs) were formulated.

RQ1 Does watching the non-storified and storified instructional videos differ in terms of changes in situational interest and positive activating epistemic emotions during the four-lesson intervention?

RQ2 Does the non-storified and storified instruction differ in terms of learning online credibility evaluation skills?

Regarding RQ1, we drew on an emotional design perspective (Mahler & Mayer, 2024; Wong & Adesope, 2021) when formulating the hypotheses. We expected that non-storified and storified conditions would differ in terms of how students' situational interest and positive activating epistemic emotions changed during the video-based instruction **(Hypothesis 1)**. We also expected that the storified videos would maintain students' situational interest and positive epistemic emotions during the intervention **(Hypothesis 1a)**, whereas, while watching the non-storified videos, students' situational interest and positive activating epistemic emotions would decline during the intervention **(Hypothesis 1b)**.

Regarding RQ2, we formulated a hypothesis and counter hypothesis for learning gains because storification can either support learning by increasing generative processing or hinder learning by increasing too much extraneous processing. If storification serves mainly as a form of emotional design that motivates students to engage in generative cognitive processing (Mahler & Mayer, 2024; Plass & Kaplan, 2016), then students in the storified condition should learn more from the intervention than those in the non-storified condition **(Hypothesis 2).** In contrast, if storification serves mainly as a form of seductive detail that distracts students and thereby increases extraneous cognitive processing (Sundararajan & Adesope, 2020; Sweller, 1988), then students in the storified condition should learn less from the intervention than those in the non-storified condition **(Alternative Hypothesis 2).**

When testing the hypotheses, we controlled for reading fluency, gender (RQ1 and RQ2), pre-test score, and the topic order of the tasks (RQ2). Reading fluency was controlled for because basic reading skills have been shown to affect reading enjoyment (Cheema, 2018; van Bergen et al., 2023) and contribute to the performance of online credibility evaluation (Hämäläinen et al., 2021; Potocki et al., 2020). Also, gender differences have been observed in reading engagement (Brozo et al., 2014; Logan & Johnston, 2009), multimedia learning (Dousay & Trujillo, 2019; Heo & Toomey, 2020), and credibility evaluation performance (Forzani, 2018; Kanniainen et al., 2019). Finally, the topic order was controlled for as the topic has been shown to be associated with the credibility evaluation performance (Bråten et al., 2018; Hämäläinen et al., 2021).

4. Method

4.1. Participants

Participants comprised 250 Finnish sixth-grade students (M_{age} = 12.77 years, SD_{age} = 0.38, girls 50.7%) representing 13 classes from 10 Finnish elementary schools. The schools were located in urban, suburban, and rural areas in five different municipalities. The majority of students (89.1%) spoke Finnish at home; 5.6% spoke Finnish and some other language at home, while 5.2% did not speak Finnish at home.

All students participated in the intervention since its content was in accordance with the Finnish national curriculum (The Finnish National Board of Education, 2016). The national curriculum includes seven areas of transversal competencies, including multiliteracies, teaching of which should be incorporated into all school subjects. Multiliteracies include abilities to seek, interpret, evaluate, and produce various types of texts in different situations. Teachers support students' multiliteracy competencies by using different media and technologies. Furthermore, the language arts curriculum includes objectives that involve guiding students in information seeking, using diverse information resources, and evaluating the credibility of information. The national curriculum forms a foundation, leaving municipalities, schools, and teachers with high autonomy for interpreting and implementing the curriculum (Palsa & Mertala, 2019).

The ethical statement was received from the Ethics Committee of the Tampere Region, and research permissions were received from the municipalities. In the study, we only used data from students who and whose guardians provided informed consent.

Thirteen volunteer classroom teachers implemented the intervention as part of their regular teaching. Of the 13 classes, six followed the non-storified intervention instruction (n = 135) and seven followed the storified intervention instruction (n = 114). Of the participants, one did not participate in any of the intervention lessons.

4.2. Materials

4.2.1. Instructional videos

In the design of instructional videos, we implemented four well-established multimedia design principles: segmentation, signaling, personalization, and positivity (Lawson et al., 2021b; Mayer, 2014). *Segmentation* was implemented by sequencing the learning content into subsets. In our case, each of the three strategy instruction videos focused on one credibility evaluation strategy (evaluation of author's expertise, author's intention, and quality of evidence) followed by a video

introducing the case to solve. The second principle, *signaling*, was used to draw students' attention to essential learning content (Mautone & Mayer, 2001; Schneider et al., 2018). The credibility evaluation strategies were highlighted by written words (non-storified condition) and by displaying each strategy on the detective's phone screen (see Fig. 2). The third and fourth principles, *positivity* and *personalization* (Mayer et al., 2004), were implemented by the instructors. The instructors used conversational, encouraging language delivered with gestures and a positive tone of voice.

The learning content of the storified and non-storified videos was similar but differed in terms of storification. Fig. 2 describes the characteristics of the storified and non-storified instructional videos. Example videos are provided in supplementary materials.

The learning materials included three types of videos. Context videos introduced the intervention (introductory video for both conditions) and provided a context (detective school) for the storified condition (trailer and concluding video celebrating the completion of the detective school). Strategy instruction videos taught the three online credibility evaluation strategies. A case video introduced a problem-solving task that required students to apply the taught online credibility evaluation strategies.

Because of the storified elements, the storified instructional videos were somewhat longer than the non-storified videos. The introductory video in the non-storified condition was 1:13 min long, whereas that in the storified condition was 2:32 min long. The length of the strategy instruction videos varied from 1:56 to 3:14 min in the non-storified condition and from 2:31 to 4:07 min in the storified condition. The non-storified case video was 1:08 min long, while the storified case video was 1:51 min long. Finally, the trailer was 1:20 min long, and the celebration video was 1:50 min long (storified condition only).

Finally, as the emotions experienced by students may have reflected instructors' emotional expressions (Lawson et al., 2021a), we examined the valence of emotions displayed by the instructors. We conducted facial expression analyses with the FaceReader software (Noldus, 2024). Instructors displayed more positive than negative emotions, and there were no considerable differences in the average valence (mean of the four videos) of conditions (see Appendix A).

4.2.2. Other learning materials

Other learning materials included a workbook, task/detective passport, and badges. The *workbook* contained learning tasks for practicing the taught credibility evaluation strategies. Tasks included identification and evaluation tasks concerning the author's expertise, the author's intentions, and the quality of evidence. In addition, students were asked to justify their responses in writing. For example, in one task concerning the author's intentions, students read four short texts from which they were asked to identify whether the author's intentions were sincere, persuasive, or commercial, and to justify their responses in writing. The instructors guided students in the workbooks, where they appeared as drawn pictures (non-storified condition) or cartoon images (storified condition). Example tasks translated from Finnish are available at https://aoe.fi/#/materiaali/3785.

Students also received a *task passport* (non-storified condition) or a *detective passport* (storified condition) for use in reflecting on their learning by considering the most important issues they learned. Students earned learning *badges* (stickers) after completing each lesson, which they collected in their passports. Non-storified learning badges included pictures of colorful candies, whereas storified learning badges had a detective school theme (e.g., a magnifying glass or Funky Facts logo).

4.3. Flow of the intervention

Students in both intervention conditions participated in four lessons (4 × 45 min) implemented by their classroom teachers over 2 weeks as part of regular schoolwork. A teacher manual described the flow of the intervention lessons, including detailed instructions for all designed activities. The teachers organized the activities, facilitated paired and classroom discussions, managed time, and supported students whenever needed.

Before the intervention lessons, students in the storified condition watched the Funky Facts detective office trailer, and students in both conditions watched the introductory video. At the end of the introductory video, the instructors stated they wanted to know students' starting level as online evaluators (a pre-test) as measured by the Online Credibility Evaluation Task (see Section 4.5.2).

In the three first intervention lessons, students were taught credibility evaluation strategies (evaluation of the author's expertise, the author's intentions, and the quality of evidence), one in each lesson. The lessons followed the same pattern and included one to three learning tasks, depending on the lesson. First, students watched the instructional video, after which they reported their situational interest and epistemic emotions as stimulated by the video content (see Section 4.5.1). Next, students individually completed a workbook task to practice the online credibility evaluation strategy they had been taught. A paired discussion on the responses followed this task. After the discussions, students received feedback from the video instructor. The classroom teacher took the role of the video instructor and read aloud feedback that encouragingly modeled the effective use of evaluation strategies. The same procedure was followed with the remaining tasks.

The fourth intervention lesson began by watching the instructional video, in which instructors introduced a "real case" that students helped to solve by employing the taught credibility evaluation strategies. To solve the case, students worked in pairs and analyzed four conflicting online texts about microplastics in tap water. Each student in the pair read two texts on a computer and evaluated their credibility with an analysis form (workbook), which guided students in considering the author's expertise, the author's intentions, and the quality of evidence. Then, each student pair shared their evaluations with each other, after which they completed a joint evaluation statement. In the storified condition, students used a form that Funky Facts detectives use when serving their customers. Here, too, students received feedback from the instructors. At the end of each lesson, students reflected on their learning in their passports and selected badges.

The following week of the fourth intervention lesson, students completed the Online Credibility Evaluation Task as a post-test. To conclude the intervention, students received diplomas. In addition, students in the storified condition were congratulated by the detectives in the short closure video.

4.4. Fidelity

The fidelity of the intervention was ensured in several ways. First, the classroom teachers participated in a 2.5-h professional development session led by the researchers. The session included an introduction to the intervention's theoretical and pedagogical basis, the learning materials, and the teacher manual, including detailed lesson plans. The teachers also had an opportunity to ask questions from the researchers. For practical reasons, two teachers could not participate in the professional development session. One of the researchers met with these teachers individually.

Second, the teachers kept diaries in which they reported any exceptions to the lesson plans. They were also asked to report the degree to which they had implemented the lesson according to plan (1 = completely according to the plan, 2 = almost according to the plan, or 3 = significantly deviating from the plan). The diaries indicated that the teachers followed the plan completely or almost completely, with one exception. One teacher reported a clear deviation in the fourth lesson. Some teachers also reported having insufficient time for the activities in the fourth lesson.

Finally, the researcher observed two lessons, the first and fourth, from each classroom teacher. Thus, in total, 26 lessons were observed. The observations were systematically recorded using a checklist

(McKenna et al., 2014). In Lesson 1, 96% of the activities were successfully implemented. The corresponding percentage was lower (87%) in Lesson 4 due to insufficient time to complete all planned activities, which was consistent with the teachers' reports.

4.5. Measures

4.5.1. Measures for situational interest and epistemic emotions

After watching each instructional video, students were asked to fill out a questionnaire concerning situational interest and epistemic emotion measures (Fig. 3). In the questionnaire, the instructor first provided general guidelines for filling out the questionnaire: "You just followed a video that considered [theme of the video]. Reflect on your experience of the instructional video and its content."

Students' situational interest was measured with a single 5-point scale item, with their responses supported by smiley faces and colors (Fig. 3). Epistemic emotions were measured with a short version of the Epistemically Related Emotions Scales (Pekrun et al., 2017) consisting of seven items (one for each epistemic emotion). In this measure, enjoyment was measured with an item of excitement. Students were asked to reflect on their experience of the instructional video and its content. They were asked to consider on a 5-point scale how strongly they had felt each emotion while following the video (Fig. 3). The present study focused only on positive activating epistemic emotions (i.e., curiosity and excitement).

4.5.2. Online credibility evaluation task

Students' learning of online credibility evaluation skills was assessed before and after the intervention with the Online Credibility Evaluation Task (Kiili et al., 2023). In the task, students were asked to read and evaluate four online texts on health-related topics. One task concerned the health effects of sugar and the other health effects of chocolate. In each task, two texts were more credible (a popular science text and a popular science news text), and two were less credible (a personal blog

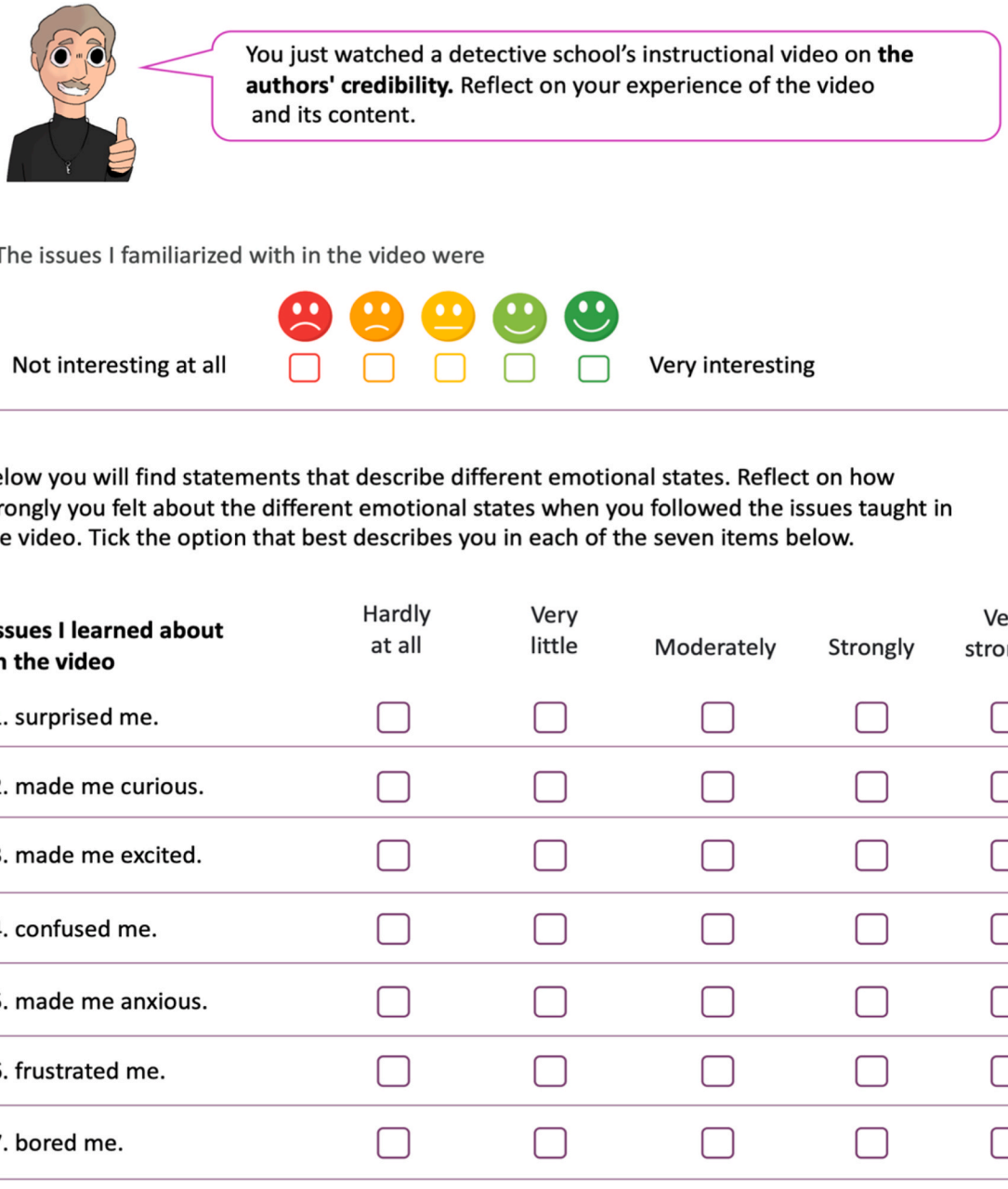


Fig. 3. Screenshot of situational interest and epistemic emotions measures included in the workbook.

text and a commercial text). Texts were manipulated in terms of the accuracy of the main claim, the author's expertise, the author's intentions, and the quality of evidence. In the more credible online texts, the main claim was in accordance with the current scientific view (e.g., chocolate does not cause acne), whereas the main claim in the less credible online texts was inaccurate (e.g., chocolate causes acne). The more credible online texts were authored by researchers or journalists who relied on expert statements and research evidence. The less credible online texts were authored by laypersons or company representatives with commercial intentions. These authors relied on either their own or customers' experiences.

Students read one text at a time. During task completion, they were guided by a fictitious fact-checker (non-storified condition) or detectives (storified condition). Students evaluated the author's expertise, the author's benevolence, and the quality of evidence in each text. After reading and evaluating each of the four texts, students evaluated overall text credibility with the following question: "How credible do you find the webpage? Please evaluate the credibility of the page on a scale from 1 to 6 [1 = "Not very credible"; 6 = "Very credible]. The higher the number, the more credible you think the page is". Students were also asked to justify their evaluations with written responses. Only overall credibility evaluations of the four online texts were used in the study. The scores for the less credible texts were reverse scored.

To counterbalance the effects of the text topic, students were randomly assigned to complete the tasks on sugar and chocolate health effects in two different orders (topic order). Topic order sugar-chocolate was coded as 0 and topic order chocolate-sugar was coded as 1. In addition, the text order within the tasks was counterbalanced. Students were randomly assigned to read the texts in one of the following orders: Order 1, Less credible text – more credible text – more credible text – less credible text; Order 2, More credible text – less credible text – less credible text – more credible text (Kiili et al., 2023).

4.5.3. Reading fluency measure

Students' reading fluency was measured with a time-limited sentence verification task from a standardized Finnish reading test battery for lower secondary school students (Lerkkanen et al., 2018). The task included 70 brief sentences (e.g., "A cow is an animal"), and students were tasked to indicate whether each sentence was correct or incorrect as rapidly as they could within 2 min. The test score was the number of correct minus incorrect responses. The task was designed so that it is not intended to reach the maximum points. Therefore, to estimate reliability, we calculated split-half reliability with Spearman-Brown correction, which was 0.97.

4.6. Data analysis

Descriptive statistics (means, standard deviations) and preliminary assessment of the similarity of the conditions in terms of reading fluency and online credibility evaluation skills (pre-test as a baseline measure) before the intervention was conducted with independent samples' *t*-tests with IBM SPSS Statistics 26 software. The results confirmed the similarity of the conditions concerning reading fluency and credibility evaluation skills before the intervention (*p*-values were all above 0.05; see Appendix B for more detailed information).

To study changes in students' situational interest, curiosity, and excitement during the intervention as well as differences in these changes across the intervention conditions (RQ1), we used latent growth curve modeling (LGCM; Muthén & Khoo, 1998) implemented in Mplus statistical software (Muthén & Muthén, 1998–2017). First, unconditional LGCMs were estimated separately for situational interest, curiosity, and excitement to examine their initial levels and subsequent changes during the intervention. For each LGCM, we estimated a level and a slope factor. The level factor reflected the initial value of an outcome at the start of the intervention, after students had watched the first strategy instruction video (Table 1). The slope factor showed the rate of change in an outcome's trajectory over time.

Table 1
Descriptives statistics of measured variables by non-storified and storified conditions.

Variable	Non-Storified			Storified		
	n	*M*	*SD*	*n*	*M*	*SD*
Situational interest at four time points (T)						
T1 Situational interest	127	3.28	0.84	103	3.24	0.83
T2 Situational interest	122	3.23	0.99	103	3.15	0.94
T3 Situational interest	114	3.37	0.86	100	3.25	0.94
T4 Situational interest	115	3.32	0.99	89	3.22	1.01
Epistemic emotions at four time points (T)						
T1 Curiosity	128	2.36	0.91	105	2.20	0.97
T2 Curiosity	126	2.13	0.89	106	2.16	1.04
T3 Curiosity	119	2.24	0.92	101	2.13	0.96
T4 Curiosity	128	2.09	0.98	97	2.01	0.96
T1 Excitement	127	2.16	0.90	105	2.17	0.96
T2 Excitement	126	2.03	0.89	105	2.04	0.96
T3 Excitement	118	1.98	0.78	101	2.03	0.96
T4 Excitement	127	2.05	0.97	97	2.05	0.96
Online credibility evaluation in pre-test						
Popular science text	134	4.49	1.13	110	4.63	1.30
Popular science news text	135	4.69	0.98	110	4.74	0.92
Personal blog text	135	3.77	1.47	111	4.09	1.33
Commercial text	134	3.81	1.41	110	3.86	1.47
Online credibility evaluation in post-test						
Popular science text	129	4.88	1.02	109	4.80	1.33
Popular science news text	129	5.04	0.85	109	4.84	1.07
Personal blog text	129	4.63	1.21	109	4.66	1.10
Commercial text	129	4.51	1.17	109	4.48	1.18
Reading fluency	131	23.32	5.21	112	23.40	7.59

Note. Scale for situational interest, curiosity, and excitement was 1–5.
Scale for online credibility items was 1–6.
Maximum score for reading fluency was 70.

To find the best-fitting model for each outcome, both linear and nonlinear shapes of the changes were explored (see Appendix C). The nonlinear change trajectory, in which the first three factor loadings were fixed as in the linear trajectory and the last loading was freely estimated (i.e., 0, 1, 2, free), best fit the data for all outcomes. The results of unconditional LGCMs are displayed in Appendix C.

Next, to examine whether the non-storified and storified instructional conditions differed in terms of change over time in situational interest, curiosity, and excitement after controlling for students' gender (0 = girl, 1 = boy) and reading fluency (as mean-centered), we conducted a conditional LGCM for each outcome. To do so, instructional conditions (non-storified = 0, storified = 1) and the two control variables were set to predict the level and slope of an outcome.

The dataset used in the analyses of RQ1 included 6.8–8.0% of missing values at T1 (Time point 1), 7.2–10.0% at T2, 12.0–14.4% at T3, and 10.0–18.4% at T4. According to Little's (1988) missing completely at random (MCAR) test, the data were not MCAR ($\chi^2(298) = 355.68$, $p = 0.012$). Yet, the missing data can be considered to be missing at random (MAR), as the reasons for missing data were tracked to students' absence from school on the day of data collection or to students skipping single items. However, as we found that boys skipped more items than girls, gender was controlled for in the analyses to strengthen the plausibility of the MAR assumption.

Following this, the full information maximum likelihood estimation was used in all LGCMs (RQ1) since it can handle all information in the data without imputing missing values (Enders, 2022). As we collected the data from 13 different classes (intra-class correlations ranged between 0.00 and 0.06), the nesting effect was also considered by implementing the TYPE = COMPLEX specification in Mplus to estimate unbiased standard errors (Muthén & Muthén, 1998–2017). The overall goodness-of-fit of all estimated LGCMs was evaluated with the following criteria (Hu & Bentler, 1999): the χ^2 test (*p*-value should be statistically non-significant), Root Mean Square Error of Approximation (*RMSEA*; the value should be < 0.06), and Standardized Root Mean

Square Residual (*SRMR*; should be < 0.08), comparative fit index, and Tucker–Lewis index (*CFI* and *TLI*, respectively; both should be > 0.95).

Regarding RQ2, we first used paired samples *t*-tests to examine whether the intervention improved all students' credibility evaluation skills on average irrespective of their intervention condition. We conducted the analysis separately for each text that students evaluated because their evaluations seemed to be somewhat unique depending on the text genre (Kiili et al., 2023). Next, to study whether conditions (non-storified = 0, storified = 1) differed in terms of students' learning during the intervention, we conducted an analysis of covariance (ANCOVA) separately for each of the evaluated texts in the online credibility evaluation task. The nesting effect of the class was controlled for by calculating a deviance score for each credibility evaluation pre- and post-test score (one pre- and post-test score per text) by subtracting a student's individual score from the mean of the corresponding variable of the student's class, thus resulting in four pre-test and four post-test deviance scores. In each ANCOVA, the dependent variable was a post-test deviance score of the credibility evaluation, and the intervention condition served as the independent variable. The corresponding pre-test online credibility evaluation deviance score, reading fluency, gender, and topic order of the task were controlled for.

5. Results

5.1. Descriptive results

The descriptive statistics for all examined variables are presented in Table 1. In both intervention conditions, students reported their situational interest to be somewhat higher, on average, than their curiosity and excitement. The Pearson correlations between the examined variables are presented in Appendix D.

5.2. Students' situational interest and positive activating epistemic emotions across the intervention conditions (RQ1)

First, we studied whether the non-storified and storified instructional conditions differed in terms of change in situational interest, curiosity, and excitement after watching an instructional video at the beginning of each of the four lessons. The results of the conditional LCGMs revealed that the intervention condition did not statistically significantly predict either the level of or change in situational interest, curiosity, and excitement (see Table 2). This suggests that the initial level and rate of change in situational interest and positive activating epistemic emotions were similar among all students, regardless of which intervention condition (non-storified or storified) they followed.

However, an examination across the intervention conditions showed that the slope mean for both curiosity and excitement was statistically significant, thus indicating that, on average, students' curiosity (Slope $M = -0.10$, $p < 0.001$) and excitement (Slope $M = -0.09$, $p < 0.001$) declined over the study period during which the intervention was implemented. Again, no statistically significant change in students' situational interest during the intervention was found.

Table 2
Standardized estimates and model fit indices for intercepts and slopes regressed on covariates (n = 241).

	Situational Interest	Curiosity	Excitement
	β	*β*	*β*
Intercept (initial level)			
Intervention condition	−0.02	−0.07	−0.00
Reading fluency	−0.06	0.08	0.02
Gender	−0.18*	−0.13	−0.22**
Slope (rate of change)			
Intervention condition	−0.08	0.04	0.04
Reading fluency	0.15*	−0.17	−0.00
Gender	0.11	0.12	0.38**
Model Fit Indexes	$\chi^2(10) = 15.70$, $p = 0.109$ *CFI* = 0.98 *TLI* = 0.97 *RMSEA* = 0.05 *SRMR* = 0.04	$\chi^2(10) = 17.36$, $p = 0.067$ *CFI* = 0.97 *TLI* = 0.94 *RMSEA* = 0.06 *SRMR* = 0.04	$\chi^2(10) = 6.39$, $p = 0.782$ *CFI* = 1.00 *TLI* = 1.00 *RMSEA* = 0.00 *SRMR* = 0.03
R^2			
Intercept	0.04	0.03	0.05
Slope	0.04	0.05	0.15

Note. $^*p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$.
For reading fluency, we used centralized mean value in the analyses.

5.3. Learning of credibility evaluation skills (RQ2)

Our second RQ examined whether the intervention developed students' credibility evaluation skills. As shown in Table 3, the intervention promoted students' credibility evaluation skills regarding the more credible texts (popular science text, popular science news text) and less credible texts (personal blog text and commercial text).

Moreover, the results showed that the intervention condition did not have any effect on the post-test score of credibility evaluations of popular science text, popular science news text, personal blog text, or commercial text (Table 4).

6. Discussion

This study investigated whether storifying instruction, and especially instructional videos, helped sixth-grade students maintain situational interest and positive activating epistemic emotions (i.e., curiosity and excitement) during a four-lesson intervention on online credibility evaluation. We also examined whether storifying instruction promoted the learning of online credibility evaluation skills. The study offers unique contributions to the design of instructions targeted to promoting students' online credibility evaluation skills. First, in the instructional design, we considered both cognitive and affective (i.e., motivational and emotional) aspects, whereas previous online credibility evaluation studies mainly focused on cognitive aspects (e.g., Braasch et al., 2013; Bråten et al., 2019; Pérez et al., 2018). Second, the study tested the usefulness of storification as a novel emotional design feature. Third, we measured situational interest and epistemic emotions during four lessons implemented in the two-week period, generating new knowledge on how situational interest, curiosity, and excitement fluctuate in an ecologically valid classroom setting.

6.1. Motivational and emotional outcomes

Regarding situational interest, the non-storified and storified conditions did not differ in terms of how students' situational interest changed during the four lessons when it was measured after they had watched either the non-storified or storified instructional video at the

Table 3
Differences in means of pre- and post-test in online credibility evaluation items across the conditions.

	Pre-test		Post-test			
Online text	*M*	*SD*	*M*	*SD*	Paired *t*-test	*d*
Popular science text	4.55	1.21	4.85	1.44	$t(234) = -3.13$, $p = 0.002$	1.44
Popular science news text	4.71	0.95	4.94	0.96	$t(235) = -3.22$, $p = 0.001$	1.13
Personal blog text	3.92	1.42	4.64	1.16	$t(236) = -7.89$, $p < 0.001$	1.42
Commercial text	3.83	1.44	4.49	1.17	$t(234) = -7.71$, $p < 0.001$	1.30

Note. Scale was 1–6.

Table 4
Differences in deviance score means of online credibility evaluation items in post-test among non-storified and storified conditions.

	Non-storified		Storified					
Online text	*M*	*SD*	*M*	*SD*	*F*	*df1, df2*	*p*	η_p^2
Popular science text	−0. 01	0.91	0.05	1.28	0.15	1, 222	0.695	0.001
Popular science news text	−0.02	0.76	0.02	1.00	0.34	1, 223	0.560	0.002
Personal blog text	−0.01	1.19	0.01	1.07	0.05	1, 224	0.817	0.000
Commercial text	0.02	1.13	0.04	1.12	0.16	1, 222	0.687	0.001

**Note*. The deviance score was calculated by subtracting a student's individual online credibility evaluation score from the mean of the corresponding variable of the students' class.

beginning of each lesson. This finding was contrary to our expectation (Hypothesis 1) because we assumed that the storified videos would maintain students' situational interest, whereas non-storified videos were expected to decline students' situational interest during the intervention. However, our findings suggest that both non-storified and storified videos successfully maintained situational interest, with no decline in situational interest being observed.

This finding could be explained by professionally produced videos in which four well-established multimedia design principles were applied (Mayer, 2014). For example, in both types of videos, instructors used a conversational style and addressed students while speaking. As the instructors in both non-storified and storified videos displayed more positive than negative emotions, students may have perceived both videos as engaging and credible (cf. Lawson et al., 2021a). Beyond this, storifying videos did not seem to confer any additional value. As the emotional design tends to be effective, and may especially motivate younger students (Brom et al., 2018), it would be worth investigating whether storifying instructional videos would be more beneficial in triggering lower primary school students' situational interest.

Notably, on average, students in the non-storified and storified conditions rated their situational interest similarly after watching the first strategy instruction video (non-storified condition, 3.28 out of 5; storified condition, 3.24 out of 5). Considering the school context, both types of videos were able to trigger considerably high situational interest. Being able to trigger and maintain students' situational interest in reading-related instruction is promising as reading motivation tends to decline in early adolescence (Miyamoto et al., 2020).

Regarding positive activating epistemic emotions, students in the non-storified and storified intervention conditions did not differ in terms of how their curiosity or excitement changed after watching instructional videos during the instruction. Again, this contradicted our expectations (Hypothesis 1). The results showed that students in both instruction conditions reported their positive activating epistemic emotions in a slightly declining manner, whereas we expected a decline only in the non-storified condition. Notably, although the decline was statistically significant, it was so small that it cannot be considered pedagogically meaningful.

It is noteworthy that students in the non-storified and storified conditions did not feel particularly curious or excited. Students rated, on average, their curiosity and excitement considerably lower than their situational interest. As curiosity requires a desire to acquire information that closes a knowledge gap (Loewenstein, 1994), the overconfidence in one's online credibility evaluation skills that we have previously observed among the same age group (Anttonen et al., 2023) may have served as one barrier to feeling curious. Aligned with this finding, a study by Metcalfe et al. (2023) found that curiosity is minimal when people are completely certain that they know the answer to a question at hand, whereas curiosity is maximal when people believe they almost know the answer. Metcalfe and colleagues also demonstrated that people become more curious when their certainty that they are correct is undermined. That is, in our case, students might have needed feedback to be able to assess their knowledge. Finally, our intervention concentrated on strategic knowledge, and knowledge gaps might be more difficult to identify compared to factual knowledge, especially when the videos did not provide any feedback or reflection prompts.

In addition, students may have watched the instructional videos rather passively. Videos without interactive elements or prompts that activate students' thinking may not effectively induce positive activating epistemic emotions. In experimental psychology, trivia or knowledge questions are commonly used to induce epistemic emotions, especially curiosity (Ozono et al., 2021). Regarding epistemic enjoyment (in our study, excitement), learners can experience it when they, for example, have satisfied their curiosity (Chevrier et al., 2019) or managed to solve problems they encounter (Pekrun & Linnenbrink-Garcia, 2012). Thus, as the videos did not induce much curiosity or include problem-solving activities, it is not surprising that students did not feel particularly excited when watching the videos. Therefore, future research could investigate how incorporating interactive elements into storified instructional videos can enhance students' positive activating epistemic emotions.

Future research should also examine effects of different storification themes and elements to understand better which kinds of frame stories or fantasies would be appealing to students of varying ages and backgrounds (Malone, 1981). Further, analyzing the data at the condition level may have concealed individual variations in student engagement. Person-centered approaches are necessary to gain a deeper understanding of how individual differences in students' prior knowledge and interests related to the learning content are associated with their motivational, emotional, and cognitive engagement.

6.2. Learning outcomes

Regarding students' learning of online credibility evaluation skills, the results showed that the non-storified and storified instructions similarly promoted students' online credibility evaluation skills. Contrary to our expectations (Hypothesis 2, Alternative Hypothesis 2), students in both conditions performed similarly in the post-test when evaluating the more credible (popular science and popular science news texts) and less credible online texts (i.e., personal blog and commercial text).

As the storified videos did not seem to motivate or emotionally engage students differently than the non-storified videos, it is not surprising that students in the storified condition did not learn more than those in the non-storified condition. Thus, contrary to our expectations, storification did not seem to increase generative processing. However, incorporating storified elements into the videos did not seem to hamper students' learning either. Thus, our study suggests that storified elements were not so seductive that they hindered learning, or at least that the possible negative effects were smaller than could be expected in light of the cognitive theory of multimedia learning (Mayer, 2014).

Notably, the instructional design of the non-storified and storified interventions was guided by several well-established multimedia learning principles (i.e., segmentation, signaling, personalization, and positivity). Consequently, the successful application of these principles may have kept the essential and extraneous cognitive load so low that the added extraneous load caused by storification did not overload learners' cognitive systems. This may partly explain why storification elements also did not undermine learning outcomes, as our alternative

hypothesis expected.

Our study did not find evidence to establish storification as a new emotional design feature. However, it is possible that storification might work in other learning domains, in different kinds of learning materials, or in different populations. In fact, Mayer (2024, p. 19) emphasized that each design principle "is subject to boundary conditions including for whom the principle applies, for which kind of lesson the principle applies, and under what circumstances the principle applies." Therefore, researchers should test storification in different settings and reveal possible boundary conditions.

It is worth highlighting that students learned to question the overall credibility of the personal blog text and commercial text during the intervention. The considerably large effect size (Cohen, 1992), especially considering the learning of questioning, is promising since previous studies have shown that students struggle more in evaluating less credible texts than more credible texts (Anttonen et al., 2023; Kiili et al., 2018).

6.3. Limitations and future research

The present study had several limitations that must be acknowledged. First, we measured situational interest and epistemic emotions only after students had watched the first strategy instruction video. Thus, we could not evaluate to what extent the introductory videos and trailer triggered students' situational interest in the first place. Since the initial level of situational interest contributes to subsequent levels of situational interest (Hidi & Renninger, 2006; Renninger et al., 2019), measuring situational interest in the beginning would have provided a more complete and comprehensive understanding of fluctuations in situational interest during the intervention.

Second, we relied solely on students' self-reports of their situational interest and epistemic emotions. Students may have filled in the measures without thoroughly considering their affective and emotional states. Also, differentiation of epistemic emotions may have been not only a novel activity for students but also a challenge in the regular classroom context. Future research could supplement self-reports with other measures, such as physiological measures (Graesser, 2020) or emote-aloud techniques (Craig et al., 2008).

The third limitation was the unequal number of instructors in the videos. The non-storified videos had one instructor, whereas the storified videos had two. Future studies could compare non-storified and storified videos with the same number of instructors. Fourth, the storified videos were longer than the non-storified ones, providing students in the storified condition with more processing time. However, all the videos were short and contained the same learning content. The additional length of the storified videos was attributed to incorporating the storified elements.

Finally, the absence of a delayed post-test limited the examination of potential storification effects on retention or transfer tests. Future research should incorporate delayed assessments to explore the long-term impact of storification as an emotional design element on learning outcomes.

7. Conclusions

This study employed a novel approach to emotional design by storifying instructional videos with on-screen human instructors. In addition, storification was extended to other learning materials to offer students a holistic learning experience. To investigate the added value of the storified instructional videos, we designed two four-lesson interventions, in which students watched either non-storified or storified instructional videos and completed related learning assignments aimed at promoting online credibility evaluation skills.

Storified instructional videos did not confer any added value in terms of learners' situational interest or positive activating epistemic emotions. Although instructional videos, regardless of their type, did maintain students' situational interest, a slight decline in positive activating epistemic emotions was observed. Finally, the non-storified and storified instructions improved students' online credibility evaluation skills similarly. These results underscore that professionally designed, segmented, and personalized videos that signal the key concepts and deliver the instruction in a positive tone can both trigger and maintain situational interest and enhance students' online credibility evaluation skills without the need to incorporate additional storified elements. Although we did not find evidence to establish storification as a new emotional design feature, our study was the first of its kind, and more research is thus needed in different settings and with different instructional content.

CRediT authorship contribution statement

Riikka Anttonen: Writing – review & editing, Writing – original draft, Resources, Methodology, Investigation, Formal analysis, Conceptualization. **Kristian Kiili:** Writing – review & editing, Supervision, Methodology, Funding acquisition, Conceptualization. **Eija Räikkönen:** Writing – review & editing, Supervision, Methodology, Formal analysis. **Carita Kiili:** Writing – review & editing, Supervision, Resources, Project administration, Methodology, Funding acquisition, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

Acknowledgments

This study was supported by the Research Council of Finland (No. 324524) and the Strategic Research Council (No. 335625, 358250). The authors would like to thank Hannu Aaltonen from Hehkumedia for the video production and actors Iina Hautala and Aleksi Kauppinen. We also thank Jemina Vahtolammi for the illustrations in the workbook and Essi Viertola for helping us in data collection.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.chb.2024.108385.

References

Alemdag, E. (2022). Effects of instructor-present videos on learning, cognitive load, motivation, and social presence: A meta-analysis. *Education and Information Technologies, 27*(9), 12713–12742. https://doi.org/10.1007/s10639-022-11154-w

Anttonen, R., Räikkönen, E., Kiili, K., & Kiili, C. (2023). Sixth graders evaluating online texts: Self-efficacy beliefs predict confirming but not questioning the credibility. *Scandinavian Journal of Educational Research.* https://doi.org/10.1080/00313831.2023.2228834. Advance online publication.

Barzilai, S., Mor-Hagani, S., Abed, F., Tal-Savir, D., Goldik, N., Talmon, I., & Davidow, O. (2023). Misinformation Is Contagious: Middle school students learn how to evaluate and share information responsibly through a digital game. *Computers & Education, 202*, Article 104832. https://doi.org/10.1016/j.compedu.2023.104832

Barzilai, S., Thomm, E., & Shlomi-Elooz, T. (2020). Dealing with disagreement: The roles of topic familiarity and disagreement explanation in evaluation of conflicting expert claims and sources. *Learning and Instruction, 69*, Article 101367. https://doi.org/10.1016/j.learninstruc.2020.101367

Barzilai, S., & Weinstock, M. (2020). Beyond trustworthiness: Comprehending multiple source perspectives. In A. List, P. Van Meter, D. Lombardi, & P. Kendeou (Eds.), *Handbook of learning from multiple representations and perspectives* (pp. 123–140). Routledge.

Beege, M., Nebel, S., Schneider, S., & Rey, G. D. (2019). Social entities in educational videos: Combining the effects of addressing and professionalism. *Computers in Human Behavior, 93*, 40–52. https://doi.org/10.1016/j.chb.2018.11.051

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

Braasch, J. L. G., & Graesser, A. C. (2019). Avoiding and overcoming misinformation on the Internet. In R. J. Sternberg, & D. F. Halpern (Eds.), *Critical thinking in psychology* (2nd ed., pp. 125–151). Cambridge University Press.

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

Bråten, I., McCrudden, M. T., Stang Lund, E., Brante, E. W., & Strømsø, H. I. (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*(3), 345–365. https://doi.org/10.1002/rrq.197

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

Brom, C., Starkova, T., & D'Mello, S. K. (2018). How effective is emotional design? A meta-analysis on facial anthropomorphisms and pleasant colors during multimedia learning. *Educational Research Review, 25*, 100–119. https://doi.org/10.1016/j.edurev.2018.09.004

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

Cheema, J. R. (2018). Adolescents' enjoyment of reading as a predictor of reading achievement: New evidence from a cross-country survey. *Journal of Research in Reading, 41*(1), 149–162. https://doi.org/10.1111/1467-9817.12257

Chevrier, M., Muis, K. R., Trevors, G. J., Pekrun, R., & Sinatra, G. M. (2019). Exploring the antecedents and consequences of epistemic emotions. *Learning and Instruction, 63*, Article 101209. https://doi.org/10.1016/j.learninstruc.2019.05.006

Chinn, C. A., Barzilai, S., & Duncan, R. G. (2021). Education for a "Post-Truth" world: New directions for research and practice. *Educational Researcher, 50*(1), 51–60. https://doi.org/10.3102/0013189X20940683

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

Craig, S. D., D'Mello, S., Witherspoon, A., & Graesser, A. (2008). Emote aloud during learning with AutoTutor: Applying the facial action coding system to cognitive–affective states during learning. *Cognition & Emotion, 22*(5), 777–788. https://doi.org/10.1080/02699930701516759

D'Mello, S. K., & Graesser, A. C. (2012). Dynamics of affective states during complex learning. *Learning and Instruction, 22*, 145–157. https://doi.org/10.1016/j.learninstruc.2011.10.001

Dousay, T. A., & Trujillo, N. P. (2019). An examination of gender and situational interest in multimedia learning environments. *British Journal of Educational Technology, 50*(2), 876–887. https://doi.org/10.1111/bjet.12610

Elliot, A. J. (2023). Energization and direction are both essential parts of motivation. In M. Bong, J. Reeve, & S. Kim (Eds.), *Motivation science: Controversies and insights* (pp. 10–14). Oxford University Press.

Enders, C. K. (2022). *Applied missing data analysis*. Guilford Publications.

Fiorella, L., & Mayer, R. E. (2018). What works and doesn't work with instructional video. *Computers in Human Behavior, 89*, 465–470. https://doi.org/10.1016/j.chb.2018.07.015

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

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

Fulmer, S. M., & Tulis, M. (2013). Changes in interest and affect during a difficult reading task: Relationships with perceived difficulty and reading fluency. *Learning and Instruction, 27*, 11–20. https://doi.org/10.1016/j.learninstruc.2013.02.001

Fyfield, M., Henderson, M., & Phillips, M. (2022). Improving instructional video design: A systematic review. *Australasian Journal of Educational Technology, 38*(3), 155–183. https://doi.org/10.14742/ajet.7296

Graesser, A. C. (2020). Emotions are the experiential glue of learning environments in the 21st century. *Learning and Instruction, 70*, Article 101212. https://doi.org/10.1016/j.learninstruc.2019.05.009

Hämäläinen, E. (2023). Examining and enhancing adolescents' critical online reading skills. *JYU Dissertations, 663*. http://urn.fi/URN:ISBN:978-951-39-9654-3.

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

Hämäläinen, E., Kiili, C., Räikkönen, E., & Marttunen, M. (2021). Students' abilities to evaluate the credibility of online texts: The role of internet-specific epistemic justifications. *Journal of Computer Assisted Learning, 37*(5), 1409–1422. https://doi.org/10.1111/jcal.12580

Harackiewicz, J., Barron, K., Tauer, J., Carter, S., & Elliot, A. (2000). Short-term and long-term consequences of achievement: Predicting interest and performance over time. *Journal of Educational Psychology, 92*, 316–330. https://doi.org/10.1037/0022-0663.92.2.316

Henderson, M. L., & Schroeder, N. L. (2021). A systematic review of instructor presence in instructional videos: Effects on learning and affect. *Computers and Education Open, 2*, Article 100059. https://doi.org/10.1016/j.caeo.2021.100059

Hendriks, F., Kienhues, D., & Bromme, R. (2015). Measuring laypeople's trust in experts in a digital age: The Muenster Epistemic Trustworthiness Inventory (METI). *PLoS One, 10*(10), Article e0139309. https://doi.org/10.1371/journal.pone.0139309

Hendriks, F., Kienhues, D., & Bromme, R. (2016). Evoking vigilance: Would you (dis) trust a scientist who discusses ethical implications of research in a science blog? *Public Understanding of Science, 25*(8), 992–1008. https://doi.org/10.1177/0963662516646048

Heo, M., & Toomey, N. (2020). Learning with multimedia: The effects of gender, type of multimedia learning resources, and spatial ability. *Computers & Education, 146*, Article 103747. https://doi.org/10.1016/j.compedu.2019.103747

Hidi, S. (2001). Interest, reading, and learning: Theoretical and practical considerations. *Educational Psychology Review, 13*, 191–209. https://doi.org/10.1023/A:1016667621114

Hidi, S., & Renninger, K. (2006). The four-phase model of interest development. *Educational Psychologist, 41*(2), 111–127. https://doi.org/10.1207/s15326985ep4102_4

Hidi, S. E., & Renninger, K. A. (2020). On educating, curiosity, and interest development. *Current Opinion in Behavioral Sciences, 35*, 99–103. https://doi.org/10.1016/j.cobeha.2020.08.002

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

Iordanou, K., Muis, K. R., & Kendeou, P. (2019). Epistemic perspective and online epistemic processing of evidence: Developmental and domain differences. *The Journal of Experimental Education, 87*(4), 531–551. https://doi.org/10.1080/00220973.2018.1482857

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

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

Kiili, C., Räikkönen, E., Bråten, I., Strømsø, H. I., & Hagerman, M. S. (2023). Examining the structure of credibility evaluation when sixth graders read online texts. *Journal of Computer Assisted Learning, 39*(3), 954–969. https://doi.org/10.1111/jcal.12779

Kiili, C., Kiili, K., Räikkönen, E., & Coiro, J. (under review). Explicit video-based instruction enhanced students' online credibility evaluation skills: Does storifying instruction matter?.

Kirkpatrick, L. C., & Klein, P. D. (2016). High-achieving high school students' strategies for writing from Internet-based sources of information. *Journal of Writing Research, 8*(1), 1–47. https://doi.org/10.17239/jowr-2016.08.01.01

Koskinen, A., McMullen, J., Hannula-Sormunen, M., Ninaus, M., & Kiili, K. (2023). The strength and direction of the difficulty adaptation affect situational interest in game-based learning. *Computers & Education, 194*, Article 104694. https://doi.org/10.1016/j.compedu.2022.104694

Kuutila, M., Kiili, C., Kupiainen, R., Huusko, E., Li, J., Hosio, S., Mäntylä, M., Coiro, J., & Kiili, K. (2024). Revealing complexities when adult readers engage in the credibility evaluation of social media posts. *Computers in Human Behavior*. , Article 108017. https://doi.org/10.1016/j.chb.2023.108017

Lawson, A. P., Mayer, R. E., Adamo-Villani, N., Benes, B., Lei, X., & Cheng, J. (2021a). Do learners recognize and relate to the emotions displayed by virtual instructors? *International Journal of Artificial Intelligence in Education, 31*, 134–153. https://doi.org/10.1007/s40593-021-00238-2

Lawson, A. P., Mayer, R. E., Adamo-Villani, N., Benes, B., Lei, X., & Cheng, J. (2021b). The positivity principle: Do positive instructors improve learning from video lectures? *Educational Technology Research & Development, 69*, 3101–3129. https://doi.org/10.1007/s11423-021-10057-w

Lerkkanen, M. K., Eklund, K., Löytynoja, H., Aro, M., & Poikkeus, A. M. (2018). *YKÄ—luku-ja kirjoitustaidon arviointimenetelmiä yläkouluun. [YKÄ - reading and writing assessments for secondary school]*. Niilo Mäki Instituutti.

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). International Reading Association.

List, A. (2024). The limits of reasoning: Students' evaluations of anecdotal, descriptive, correlational, and causal evidence. *The Journal of Experimental Education, 92*(1), 1–31. https://doi.org/10.1080/00220973.2023.2174487

Little, R. J. (1988). A test of missing completely at random for multivariate data with missing values. *Journal of the American Statistical Association, 83*(404), 1198–1202. https://doi.org/10.1080/01621459.1988.10478722

Loderer, K., Pekrun, R., & Lester, J. C. (2020). Beyond cold technology: A systematic review and meta-analysis on emotions in technology-based learning environments. *Learning and Instruction, 70*, Article 101162. https://doi.org/10.1016/j.learninstruc.2018.08.002

Loewenstein, G. (1994). The psychology of curiosity: A review and reinterpretation. *Psychological Bulletin, 116*(1), 75–98. https://doi.org/10.1037/0033-2909.116.1.75

Logan, S., & Johnston, R. (2009). Gender differences in reading ability and attitudes: Examining where these differences lie. *Journal of Research in Reading, 32*(2), 199–214. https://doi.org/10.1111/j.1467-9817.2008.01389.x

Logtenberg, A., Van Boxtel, C., & van Hout-Wolters, B. (2011). Stimulating situational interest and student questioning through three types of historical introductory texts. *European Journal of Psychology of Education, 26*, 179–198. https://doi.org/10.1007/s10212-010-0041-6

Lucassen, T., Muilwijk, R., Noordzij, M. L., & Schraagen, J. M. (2013). Topic familiarity and information skills in online credibility evaluation. *Journal of the American Society for Information Science and Technology, 64*(2), 254–264. https://doi.org/10.1002/asi.22743

Magner, U. I., Schwonke, R., Aleven, V., Popescu, O., & Renkl, A. (2014). Triggering situational interest by decorative illustrations both fosters and hinders learning in computer-based learning environments. *Learning and Instruction, 29*, 141–152. https://doi.org/10.1016/j.learninstruc.2012.07.002

Mahler, L. S., & Mayer, R. E. (2024). Anime as a medium for science learning. *Journal of Computer Assisted Learning, 40*(2), 787–796. https://doi.org/10.1111/jcal.12908

Malone, T. W. (1981). Toward a theory of intrinsically motivating instruction. *Cognitive Science, 5*(4), 333–369. https://doi.org/10.1016/S0364-0213(81)80017-1

Mautone, P. D., & Mayer, R. E. (2001). Signaling as a cognitive guide in multimedia learning. *Journal of Educational Psychology, 93*, 377–389. https://doi.org/10.1037/0022-0663.93.2.377

Mayer, R. E. (2014). Cognitive theory of multimedia learning. In R. E. Mayer (Ed.), *The Cambridge handbook of multimedia learning* (2nd ed., pp. 43–72). Cambridge University Press.

Mayer, R. E. (2019). Computer games in education. *Annual Review of Psychology, 70*, 531–549. https://doi.org/10.1146/annurev-psych-010418-102744

Mayer, R. E. (2020). *Multimedia learning* (3rd ed.). Cambridge University Press.

Mayer, R. E. (2021). Evidence-based principles for how to design effective instructional videos. *Journal of Applied Research in Memory and Cognition, 10*(2), 229–240. https://doi.org/10.1016/j.jarmac.2021.03.007

Mayer, R. E. (2024). The past, present and future of the cognitive theory of multimedia learning. *Educational Psychology Review, 36*(8). https://doi.org/10.1007/s10648-023-09842-1

Mayer, R. E., Fennell, S., Farmer, L., & Campbell, J. (2004). A personalization effect in multimedia learning: Students learn better when words are in conversational style rather than formal style. *Journal of Educational Psychology, 96*, 389–395. https://doi.org/10.1037/0022-0663.96.2.389

Mayer, R. E., Fiorella, L., & Stull, A. (2020). Five ways to increase the effectiveness of instructional video. *Educational Technology Research & Development, 68*(3), 837–852. https://doi.org/10.1007/s11423-020-09749-6

McGrew, S. (2021). Skipping the source and checking the contents: An in-depth look at students' approaches to web evaluation. *Computers in the Schools, 38*(2), 75–97. https://doi.org/10.1080/07380569.2021.1912541

McKenna, J. W., Flower, A., & Ciullo, S. (2014). Measuring fidelity to improve intervention effectiveness. *Intervention in School and Clinic, 50*(1), 15–21. https://doi.org/10.1177/1053451214532348

Metcalfe, J., Vuorre, M., Towner, E., & Eich, T. S. (2023). Curiosity: The effects of feedback and confidence on the desire to know. *Journal of Experimental Psychology: General, 152*(2), 464–482. https://doi.org/10.1037/xge0001284

Miyamoto, A., Murayama, K., & Lechner, C. M. (2020). The developmental trajectory of intrinsic reading motivation: Measurement invariance, group variations, and implications for reading proficiency. *Contemporary Educational Psychology, 63*, Article 101921. https://doi.org/10.1016/j.cedpsych.2020.101921

Muthén, B. O., & Khoo, S. T. (1998). Longitudinal studies of achievement growth using latent variable modeling. *Learning and Individual Differences, 10*(2), 73–101. https://doi.org/10.1016/S1041-6080(99)80135-6

Muthén, L. K., & Muthén, B. O. (1998). *Mplus user's guide* (Eight edition). Muthén & Muthén.

Noldus. (2024). *FaceReader: Tool for automatic analysis of facial expression: Version 9.0 [Software]*. Noldus Information Technology B. V.

Osborne, J., & Pimentel, D. (2022). Science, misinformation, and the role of education. *Science, 378*(6617), 246–248.

Ozono, H., Komiya, A., Kuratomi, K., Hatano, A., Fastrich, G., Raw, J. A. L., … Murayama, K. (2021). Magic curiosity arousing tricks (MagicCATs): A novel stimulus collection to induce epistemic emotions. *Behavior Research Methods, 53*, 188–215. https://doi.org/10.3758/s13428-020-01431-2

Palsa, L., & Mertala, P. (2019). Multiliteracies in local curricula: Conceptual contextualizations of transversal competence in the Finnish curricular framework. *Nordic Journal of Studies in Educational Policy, 5*(2), 114–126. https://doi.org/10.1080/20020317.2019.1635845

Park, B., Knörzer, L., Plass, J. L., & Brünken, R. (2015). Emotional design and positive emotions in multimedia learning: An eyetracking study on the use of anthropomorphisms. *Computers & Education, 86*, 30–42. https://doi.org/10.1016/j.compedu.2015.02.016

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

Pekrun, R., Goetz, T., Titz, W., & Perry, R. P. (2002). Academic emotions in students' self-regulated learning and achievement: A program of qualitative and quantitative research. *Educational Psychologist, 37*(2), 91–105. https://doi.org/10.1207/S15326985EP3702_4

Pekrun, R., & Linnenbrink-Garcia, L. (2012). Academic emotions and student engagement. In S. L. Christenson, A. L. Reschly, & C. Wylie (Eds.), *Handbook of research on student engagement* (pp. 259–282). Springer US.

Pekrun, R., & Linnenbrink-Garcia, L. (Eds.). (2014). *International handbook of emotions in education*. Routledge.

Pekrun, R., & Marsh, H. W. (2022). Research on situated motivation and emotion: Progress and open problems. *Learning and Instruction*. , Article 101664. https://doi.org/10.1016/j.learninstruc.2022.101664

Pekrun, R., Muis, K. R., Frenzel, A. C., & Götz, T. (2018). *Emotions at school*. Routledge.

Pekrun, R., Vogl, E., Muis, K. R., & Sinatra, G. M. (2017). Measuring emotions during epistemic activities: The epistemically-related emotion scales. *Cognition & Emotion, 31*(6), 1268–1276. https://doi.org/10.1080/02699931.2016.1204989

Pérez, A., Potocki, A., Stadtler, M., Macedo-Rouet, M., Paul, J., Salmerón, L., & Rouet, J. F. (2018). Fostering teenagers' assessment of information reliability: Effects of a classroom intervention focused on critical source dimensions. *Learning and Instruction, 58*, 53–64. https://doi.org/10.1016/j.learninstruc.2018.04.006

Pink, A., & Newton, P. M. (2020). Decorative animations impair recall and are a source of extraneous cognitive load. *Advances in Physiology Education, 44*(3), 376–382. https://doi.org/10.1152/advan.00102.2019

Plass, J., & Kaplan, U. (2016). Emotional design in digital media for learning. In S. Y. Tettegah, & M. M Gartmeier (Eds.), *Emotions, technology, design, and learning* (pp. 131–161). Elsevier Science & Technology.

Polat, H. (2023). Instructors' presence in instructional videos: A systematic review. *Education and Information Technologies, 28*(7), 8537–8569. https://doi.org/10.1007/s10639-022-11532-4

Potocki, A., de Pereyra, G., Ros, C., Macedo-Rouet, M., Stadtler, M., Salmerón, L., & Rouet, J. F. (2020). The development of source evaluation skills during adolescence: Exploring different levels of source processing and their relationships. *Journal for the Study of Education and Development, 43*(1), 19–59. https://doi.org/10.1080/02103702.2019.1690848

Renninger, K. A., Bachrach, J. E., & Hidi, S. E. (2019). Triggering and maintaining interest in early phases of interest development. *Learning, Culture and Social Interaction, 23*, Article 100260. https://doi.org/10.1016/j.lcsi.2018.11.007

Schneider, S., Beege, M., Nebel, S., & Rey, G. D. (2018). A meta-analysis of how signaling affects learning with media. *Educational Research Review, 23*, 1–24. https://doi.org/10.1016/j.edurev.2017.11.001

Schneider, S., Häßler, A., Habermeyer, T., Beege, M., & Rey, G. D. (2019). The more human, the higher the performance? Examining the effects of anthropomorphism on learning with media. *Journal of Educational Psychology, 111*(1), 57–72. https://doi.org/10.1037/edu0000273

Schrader, C., Seufert, T., & Zander, S. (2021). Learning from instructional videos: Learner gender does matter; speaker gender does not. *Frontiers in Psychology, 12*, Article 655720. https://doi.org/10.3389/fpsyg.2021.655720

Schunk, D. H. (2023). Cognitions and emotions energize and sustain motivation. In M. Bong, J. Reeve, & S. Kim (Eds.), *Motivation science: Controversies and insights* (pp. 10–14). Oxford University Press.

Shin, D. D., & Kim, S. I. (2019). Homo curious: Curious or interested? *Educational Psychology Review, 31*(4), 853–874. https://doi.org/10.1007/s10648-019-09497-x

Skulmowski, A., & Xu, K. M. (2022). Understanding cognitive load in digital and online learning: A new perspective on extraneous cognitive load. *Educational Psychology Review, 34*(1), 171–196. https://doi.org/10.1007/s10648-021-09624-7

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

Sundararajan, N., & Adesope, O. (2020). Keep it coherent: A meta-analysis of the seductive details effect. *Educational Psychology Review, 32*(3), 707–734. https://doi.org/10.1007/s10648-020-09522-4

Sweller, J. (1988). Cognitive load during problem solving: Effects on learning. *Cognitive Science, 12*(2), 257–285. https://doi.org/10.1016/0364-0213(88)90023-7

Sweller, J., van Merriënboer, J. J. G., & Paas, F. (2019). Cognitive architecture and instructional design: 20 years later. *Educational Psychology Review, 31*, 261–292. https://doi.org/10.1007/s10648-019-09465-5

The Finnish National Board of Education. (2016). *National core curriculum for basic education 2014*. The Finnish National Board of Education.

Um, E., Plass, J. L., Hayward, E. O., & Homer, B. D. (2012). Emotional design in multimedia learning. *Journal of Educational Psychology, 104*(2), 485–498. https://doi.org/10.1037/a0026609

van Bergen, E., Hart, S. A., Latvala, A., Vuoksimaa, E., Tolvanen, A., & Torppa, M. (2023). Literacy skills seem to fuel literacy enjoyment, rather than vice versa. *Developmental Science, 26*(3), Article e13325. https://doi.org/10.1111/desc.13325

van Strien, J., Kammerer, Y., Brand-Gruwel, S., & Boshuizen, H. (2016). How attitude strength biases information processing and evaluation on the web. *Computers in Human Behavior, 60*, 245–252. https://doi.org/10.1016/j.chb.2016.02.057

Vilhunen, E., Turkkila, M., Lavonen, J., Salmela-Aro, K., & Juuti, K. (2022). Clarifying the relation between epistemic emotions and learning by using experience sampling method and pre-posttest design. *Frontiers in Education, 7*, 826–852. https://doi.org/10.3389/feduc.2022.826852

Wang, J., Antonenko, P., Keil, A., & Dawson, K. (2020). Converging subjective and psychophysiological measures of cognitive load to study the effects of instructor-present video. *Mind, Brain, and Education, 14*(3), 279–291. https://doi.org/10.1111/mbe.12239

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*(4), 1060–1106. https://doi.org/10.3102/0002831209333183

Wong, R. M., & Adesope, O. O. (2021). Meta-analysis of emotional designs in multimedia learning: A replication and extension study. *Educational Psychology Review, 33*(2), 357–385. https://doi.org/10.1007/s10648-020-09545-x