# Distracted by a Talking Head? Effects of Talking Heads in Educational Videos on Learning Outcomes, Eye Movements, and Learners' Ratings

**Dissertation**

der Mathematisch-Naturwissenschaftlichen Fakultät
der Eberhard Karls Universität Tübingen
zur Erlangung des Grades eines
Doktors der Naturwissenschaften
(Dr. rer. nat.)

vorgelegt von
Christina Sondermann
aus Olpe

Tübingen
2024

Gedruckt mit Genehmigung der Mathematisch-Naturwissenschaftlichen Fakultät der Eberhard Karls Universität Tübingen.

| | |
|---|---|
| Tag der mündlichen Qualifikation: | 29.07.2024 |
| Dekan: | Prof. Dr. Thilo Stehle |
| 1. Berichterstatter: | Prof. Dr. Markus Huff |
| 2. Berichterstatter: | Prof. Dr. Stephan Schwan |
| 3. Berichterstatter: | Prof. Dr. Alexander Eitel |

**Danksagung**

Ich möchte mich bei einigen Personen bedanken, die mich im Rahmen dieser Arbeit in den letzten Jahren unterstützt haben. Mein allererster und besonderer Dank gilt Martin, der mich von Anfang bis Ende an auf diesem Weg begleitet hat und dessen Tür immer offen stand für alle meine Anliegen. Ich weiß deine tolle Unterstützung, die maßgeblich zum Gelingen dieser Dissertation beigetragen hat, wirklich sehr zu schätzen. Ich habe unglaublich viel von dir gelernt und immer große Freude an der gemeinsamen Arbeit gehabt.

Außerdem möchte ich mich bei Markus Huff und Stephan Schwan, die mich bei der Anfertigung dieser Dissertation ebenfalls sehr unterstützt haben, für die Begutachtung dieser Arbeit bedanken. Ein besonderer Dank gilt außerdem allen Hilfskräften (Danea Keuchel, Aisyah Nabilah Quamila, Inka Krüger, Lara Kläffling, David Rohde) und allen weiteren Personen, die an verschiedenen Stellen meine Arbeit unterstützt und ermöglicht haben.

Außerdem möchte ich mich bei all meinen lieben Kolleginnen und Kollegen am DIE bedanken, die meine Promotionszeit sehr bereichert haben. Ein besonderer Dank gilt dabei Julia und Fabian, die mir durch die Aufnahme in ihr Büro den Start in die Promotion sehr viel leichter gemacht haben. Außerdem möchte ich mich besonders bei Hadjar, Sina und Karoline bedanken, die ebenfalls immer ein offenes Ohr hatten und mit denen ich mich so oft austauschen konnte.

Darüber hinaus möchte ich mich bei all meinen Freundinnen – ob aus Bonn, Köln, Olpe oder darüber hinaus – bedanken, die mich ebenfalls schon lange – teilweise seit der Schulzeit – begleiten. Auch wenn es während der Promotionszeit nicht immer einfach war, war und ist die Zeit mit euch mir immer ein großer Rückhalt und eine tolle Ablenkung.

Mein letzter und größter Dank gilt meiner Familie und Steffen. Steffen, dir danke ich dafür, dass du mich immer bedingungslos unterstützt hast und mir so oft den Rücken freigehalten hast. Meiner Schwester Caroline möchte ich besonders dafür danken, dass sie immer an mich geglaubt hat und immer meine beste Freundin war. Meinen Eltern danke ich dafür, dass sie mich schon mein ganzes Leben lang voller Unterstützung und Liebe begleiten. Vielen Dank für euren unermüdlichen Einsatz – während der gesamten Schulzeit, des Studiums und der Promotion – von dem ich weiß, dass er nicht selbstverständlich ist. Meinem Vater möchte ich dabei besonders dafür danken, dass er in mir die Freude am (formalen, aber besonders am informellen) Lernen geweckt hat. Gleiches gilt für meine Mutter, die für mich immer und in jeder Lebenslage der größte Rückhalt und die erste Ansprechpartnerin ist. Ohne euch wäre ich nicht, wo und wer ich heute bin. *Danke!*

***

**Inhaltsverzeichnis**

## Summary

Educational videos are a central component in formal and informal learning contexts. When optimising the design of such educational videos, the question arises whether an instructor presenting the content should be visible next to the learning content. This often involves so-called talking heads, where the head of the instructor speaking into the camera is visible in one corner of the educational video. Theories such as the social agency theory and the cognitive theory of multimedia learning (e.g., Mayer, 2021) suggest both potential advantages (social presence, deeper processing) and disadvantages (distraction, cognitive load) of a visible instructor. For the talking head format in particular, it is questionable whether the potential disadvantages outweigh the advantages because talking heads attract visual attention but offer fewer advantages than an instructor's full body representation (e.g., due to the lack of gestures). Research concerning the talking head format is still limited, especially with regard to measures beyond learning outcomes.

The aim of this thesis was to expand the findings on the talking head format and at the same time to investigate whether certain characteristics of the learning material (e.g., the type of content alongside the talking head) represent boundary conditions for the use of talking heads. Four experiments with a total of $N = 488$ participants served as the empirical basis for the dissertation, which investigated the effects of talking heads in educational videos with narrated slides. The investigations included a variety of dependent variables, for instance different learning outcomes, eye movements, cognitive load, social presence, and satisfaction ratings. Three experiments were conducted online, and one experiment took place in the laboratory using eye tracking. In all experiments, the presence of a talking head in the educational videos was varied (either as a between- or within-subjects factor), with Experiments 1a, 1b, and 2 additionally varying the slide type (graphic- vs. text-based slides) and Experiments 1b and 2 additionally varying the presentation type of the content (sequential step-by-step presentation or static all at once presentation). The results largely showed no effects of the talking head on learning outcomes (Experiments 1a, 1b, and 3), although in one investigation (Experiment 2) there was a small detrimental talking head effect on learning outcomes. With regard to learners' ratings, the findings were heterogeneous, with Experiment 2 showing strong positive effects on different ratings (e.g., satisfaction), whereas such effects were mostly absent in the remaining experiments (Experiments 1a, 1b, and 3). However, at the level of process data measured with learners' eye movements, there was a strong distraction from the learning content due to the talking head (Experiment 3). To further substantiate the findings, internal

meta-analyses across the four experiments were conducted on the key dependent variables captured in all experiments. Analyses revealed that the talking head had no overall effect on learning outcomes across all experiments, but enhanced learners' ratings of satisfaction and decreased their self-reported extraneous cognitive load. Taken together, considering the findings of the internal meta-analyses, the advantages outweigh the disadvantages, given that a talking head can improve individuals' learning experience.

## Zusammenfassung

Lernvideos sind zentraler Bestandteil in formellen und informellen Lernkontexten. Bei der optimalen Gestaltung solcher Lernvideos stellt sich auch immer wieder die Frage, ob neben den Lerninhalten auch die Lehrperson, die die Inhalte präsentiert, sichtbar sein sollte. Dies geschieht häufig in Form von sogenannten Talking Heads, bei denen der Kopf der Lehrperson, die in die Kamera spricht, in einer Ecke des Lernvideos zu sehen ist. Theorien wie die Social Agency Theory und die Cognitive Theory of Multimedia Learning (z.B. Mayer, 2021) legen sowohl potenzielle Vorteile (soziale Präsenz und tiefere Verarbeitung) als auch Nachteile (Ablenkung, kognitive Belastung) einer sichtbaren Lehrperson nahe. Insbesondere für das Talking Head Format stellt sich die Frage, ob die potenziellen Nachteile gegenüber den Vorteilen überwiegen, weil Talking Heads visuelle Aufmerksamkeit auf sich ziehen, aber weniger Vorteile bieten (z. B. aufgrund fehlender Gesten) als Ganzkörperdarstellungen einer Lehrperson. Die Forschung in Bezug auf das Talking Head Format ist noch sehr begrenzt, vor allem im Hinblick auf Maße jenseits von Lernergebnissen.

Mit dieser Arbeit sollte die Befundlage zum Format Talking Head erweitert und gleichzeitig untersucht werden, ob bestimmte Eigenschaften des Lernmaterials (z.B. die Art der Inhalte, die neben dem Talking Head erscheinen) Randbedingungen für den Einsatz von Talking Heads darstellen. Als empirische Grundlage der Untersuchung dienten vier Experimente mit insgesamt $N = 488$ Teilnehmenden, wobei die Effekte von Talking Heads in Lernvideos mit Folien untersucht wurden. Die Untersuchungen umfassten eine Vielzahl von abhängigen Variablen, zum Beispiel verschiedene Lernergebnisse, Augenbewegungen, kognitive Belastung und Bewertungen der sozialen Präsenz und der Zufriedenheit. Drei Experimente wurden online durchgeführt und ein Experiment fand im Labor mit Eyetracking statt. In allen Experimenten wurde die Anwesenheit eines Talking Heads in den Lernvideos variiert (entweder als Zwischen- oder Innersubjektfaktor), wobei in den Experimenten 1a, 1b und 2 zusätzlich der Folientyp (grafik- vs. textbasierte Folien) und in den Experimenten 1b und 2 zusätzlich die Art der Präsentation des Inhalts (sequenzielle Schritt für Schritt Präsentation oder statische Präsentation auf einmal) variiert wurde. Die Ergebnisse zeigten größtenteils keine Effekte des Talking Heads auf Lernergebnisse (Experimente 1a, 1b und 3), wobei sich in Experiment 2 ein kleiner lernhinderlicher Effekt des Talking Heads ergab. Im Hinblick auf die Bewertungen der Videos durch die Lernenden waren die Ergebnisse heterogen, wobei Experiment 2 starke positive Effekte auf die Bewertungen (z.B. Zufriedenheit) zeigte, während solche Effekte in den übrigen Experimenten (Experimente 1a, 1b und 3) meist nicht auftraten.

Auf der Ebene von Prozessdaten, die mit den Blickbewegungen der Lernenden gemessen wurden, zeigte sich jedoch eine starke Ablenkung vom Lerninhalt durch den Talking Head (Experiment 3). Um die Befunde weiter zu untermauern, wurden interne Meta-Analysen über die vier Experimente bezüglich der zentralen abhängigen Variablen, die in allen Experimenten erfasst wurden, durchgeführt. Diese ergaben, dass der Talking Head über alle Experimente hinweg insgesamt keinen Einfluss auf die Lernergebnisse hatte, aber die Zufriedenheit der Lernenden steigerte und die von ihnen selbst berichtete kognitive Belastung reduzierte. Betrachtet man die Befunde der internen Meta-Analysen, so überwiegen also die Vorteile gegenüber den Nachteilen, in Anbetracht der Tatsache, dass der Talking Head das Lernerlebnis von Personen verbessern kann.

## List of Publications Included in the Dissertation

### a) Accepted Publications

Sondermann, C., & Merkt, M. (2023). What is the effect of talking heads in educational videos with different types of narrated slides? *Contemporary Educational Psychology*, *74*, Article 102207. https://doi.org/10.1016/j.cedpsych.2023.102207

Sondermann, C., & Merkt, M. (2023). Like it or learn from it: Effects of talking heads in educational videos. *Computers & Education*, *193*, Article 104675. https://doi.org/10.1016/j.compedu.2022.104675

Sondermann, C., Huff, M., & Merkt, M. (2024). Distracted by a talking head? An eye tracking study on the effects of instructor presence in learning videos with animated graphic slides. *Learning and Instruction, 91*, Article 101878. https://doi.org/10.1016/j.learninstruc.2024.101878

# 1. General Introduction

Video-based learning and online learning in general are becoming increasingly important, further driven by the COVID-19 pandemic. One question repeatedly arises when designing educational videos: Is it useful to see a talking head alongside the actual learning content? From a practical perspective, lecturers face the question of whether they should turn on their camera or not when producing online lectures or lecture slides. This dissertation aims to contribute towards answering this question. The term *talking head* refers to a format of instructor presence in educational videos in which only the instructor's head is visible next to the learning content lecturing into the camera (cf. Hansch et al., 2015). A visible talking head may act as a social cue, initiating social processes and deeper processing in online learning (e.g., Mayer, 2021). At the same time, humans show a preference for other humans' faces already in their very early childhood and tend to look at them frequently (Gullberg & Holmqvist, 2006; Johnson et al., 1991; Võ et al., 2012). This preference for human faces harbours the risk that a visible talking head in learning materials may also lead to distraction. In addition, the talking head may increase learners' cognitive load (Mayer, 2021).

In the following sections, I will first focus on visible instructors in educational videos in general. This is followed by a more precise definition of talking heads and the presentation of the theoretical and empirical foundations of this thesis. I will continue by describing the four experiments that constitute the empirical work of this thesis. This is supplemented by internal meta-analyses across the four experiments included in this thesis with regard to the central dependent variables to further substantiate the findings. Finally, I will summarise the results, relate them to each other, and discuss implications, strengths, limitations, and future directions for research.

## 1.1. Instructor Presence in Educational Videos

Educational videos are becoming increasingly important and represent a central component of both informal and formal learning contexts for a considerable time (e.g., Feierabend & Klinger, 2003; Guo et al., 2014; Hew & Lo, 2018; Rat für kulturelle Bildung, 2019; ten Hove & van der Meij, 2015; van Alten et al., 2019). Potential formal and informal learning contexts include, for instance, watching a YouTube video in order to use a certain Excel function, participating in a Massive Open Online Course (MOOC) on Business Administration, watching a video about van Gogh in the classroom as an introduction to the lesson, watching videos on direct and alternating current electric circuits at home in order to better understand the homework, or watching lecture videos on anatomy from the previous

semester in preparation for the exam. Although many different types of videos and contexts were mentioned above, a central question arises in all of them: Should an instructor be visible in the video and under what conditions?

The increasing importance of educational videos is evident in a variety of contexts. For example, 86% of U.S. users state that they use YouTube to learn new things (O'Neil-Hart, 2017). In 2022, Coursera as one of the largest online learning platforms in the world counts more than 100 million users worldwide (Coursera, 2023). The increasing use of videos and online learning is also evident for academic courses in Germany, which may also result from the COVID-19 pandemic (Skulmowski & Rey, 2020). It seems as if a lot of content that used to be presented in analogue form (e.g., lectures) is now available online as video.

Educational videos offer enormous potential because they enable learning independent of time and location (e.g., Kay, 2012). In addition, new technical developments allow many people to produce educational videos without much effort. For instance, PowerPoint now enables a direct integration of the speaker's video into a slideshow. However, when creating such educational videos, the person producing them is naturally faced with a number of decisions: How to arrange content such as corresponding words and pictures in the learning material? Should the person presenting the content be visible in the video? Should the content be presented text-based or with graphics and images? And should complex content appear on the slides step by step or all at once? Multimedia research has already provided helpful answers to some of the aforementioned questions, for instance by defining well-established principles such as the spatial contiguity principle, which states that corresponding words and pictures in the learning material should be arranged near rather than far from each other (e.g., Mayer, 2021). However, the question concerning the instructor's visibility is still unresolved and involves competing perspectives: On the one hand, the instructor as a social cue may lead to deeper processing, but on the other hand, it may also distract or increase cognitive load (e.g., Mayer, 2021). Further, it is unclear whether the effect of instructor's visibility is affected by other design decisions such the use of graphic-based content or stepwise presentation.

In this dissertation, I will focus on videos of human instructors. There is a broad spectrum of visible instructors in educational videos including also animated pedagogical agents, offering new possibilities but also challenges that exceed the scope of this thesis. For instance, pedagogical agents can differ greatly in how human-like they appear and perform (e.g., Fiorella & Mayer, 2022a). As this brings even more variance into the already heterogeneous instructor presence research, the following descriptions focus on human

instructors to ensure consistency. A corresponding focus on human instructors is also evident in recent research overviews (e.g., Alemdag, 2022; Beege et al., 2023).

Regarding educational videos including a talking human instructor, various formats can be distinguished according to different dimensions. Some of these formats and their different dimensions are listed below:

- First, formats may be characterised by which part of the instructor is visible, for instance the whole body (e.g., Hew & Lo, 2020), only the upper body (e.g., van Wermeskerken et al., 2018), only the instructor's hand (e.g., Schroeder, 2016), or only the instructor's head (e.g., Kizilcec et al., 2014). In this regard, several studies (e.g., Beege et al., 2020; Dargue et al., 2019) demonstrated that visible gestures (e.g., as referential cues) can have learning-enhancing effects.
- In addition, it may vary whether the instructor turns toward or away from the learners and whether the instructor speaks directly into the camera or not (e.g., Beege et al., 2017; Pi et al., 2020). In this respect, Beege et al. (2017) showed, that a frontal orientation (facing the camera directly, with eye contact) led to better learning outcomes than a lateral (i.e., slightly averted) orientation. Additionally, Pi et al. (2020) observed no effect of body orientation, but learning-hindering effects of an instructor's averted gaze. Besides this, eye tracking analyses revealed that an instructor's direct gaze resulted in learners paying more attention to the face than when the instructor (partially) averted her gaze (Pi et al., 2020).
- Further, the position of the instructor's image in the video (right, left, top, bottom, middle) may vary from video to video, as can the size at which the instructor appears. For instance, Zhang et al. (2022) showed that instructors appearing in the middle of a slide attracted particularly high levels of learners' attention. Additionally, in the study by Pi et al. (2017a), learners learnt better with small images of the instructor than with medium or large ones.
- In addition, the visible instructor may be integrated into learning material in different ways, e.g., including a picture-in-picture presentation (e.g., Kokoç et al., 2020; Ng and Przybylek, 2021) or a direct integration (e.g., Kizilcec et al., 2014), which can be realised, for instance, by using a green screen.
- Beyond, instructors may differ in terms of the tasks they perform: A visible instructor can demonstrate an action (e.g., van Gog et al., 2014), point to learning content with

gestures (e.g., Pi et al., 2017b), draw a diagram (e.g., Fiorella et al., 2019), or simply speak into the camera and explain something (e.g., Kizilcec et al., 2014).

As different as these formats are, it is hardly surprising that research on instructor presence reveals heterogeneous findings (e.g., Hong et al., 2018; Kizilcec et al., 2014; Wilson et al., 2018). The variety of design options also implies that a limitation and consistency in the choice of instructor presence format is necessary, leading me to focus on the talking head format for all studies included in this dissertation. The reasons for this decision are outlined in the following sections.

Since I started my PhD, some overviews of instructor presence research in general were published, including two systematic reviews (Henderson & Schroeder, 2021; Polat, 2023) and two meta-analyses (Alemdag, 2022; Beege et al., 2023). In the recent meta-analysis by Beege et al. (2023), two articles included in this thesis have already been considered and have influenced the overall results, for instance by contributing to the instructor effect on social presence reported in this meta-analysis. To avoid circularity, I do not elaborate on this meta-analysis (Beege et al., 2023) at this point when describing the underlying research. However, I will highlight its results in the course of the General Discussion section. The review by Henderson and Schroeder (2021), which included 12 studies, found no consistent evidence in favour of including or excluding an on-screen instructor in terms of learning outcomes. However, individual studies within the review suggested that an instructor may increase learners' satisfaction. The systematic review of 41 empirical studies by Polat (2023) also concluded that instructor presence in videos is not generally beneficial for learning outcomes, but specified that, for instance, attentional cues provided by an instructor may promote learning. Further, this review studied the positive effects of instructor presence on learners' affective ratings. Following the idea that the heterogeneous findings on instructor presence could result from to the instructors' heterogeneity, Alemdag (2022) attempted to systematise the findings by including different formats of instructor presence as a potential moderator in her meta-analysis. This synthesis of 20 experimental studies reported no significant effects of instructor presence on learning measures and social presence, but an increase in ratings of cognitive load and motivation. In addition, marginal effects at least suggested that different formats of instructor presence, namely the instructor's hand, only the instructor's talking head, and other types of instructors (e.g., the instructor's upper body), might have different effects, although especially the numbers of individual studies on the hand and talking head formats were rather small.

In general, the number of studies explicitly investigating the talking head format and its effect on learning outcomes was very limited when my thesis started, despite its frequent use (e.g., Chen & Wu, 2015; Kizilcec et al., 2015). I am aware of only few experiments that already existed before my first experiment in 2020 (Hew & Lo, 2020; Kizilcec et al., 2014; Kizilcec et al., 2015; Rosenthal & Walker, 2020), with only two of them involving a control condition without talking head (Kizilcec et al., 2014; Rosenthal & Walker, 2020).[1] However, numerous other studies investigating this format (Kokoç et al., 2020; Ng and Przybylek, 2021; Yuan et al., 2021; Zhang & Yang, 2022) were published in the last years (since my first experiment in 2020) in addition to my own studies. This increase in research interest may potentially also stem from the COVID-19 pandemic, which gave new relevance to topics such as digital learning, online learning, video-based learning, as well as social processes in learning, for instance in order to prevent isolation (e.g., Baber, 2022; Schneider et al., 2022). Before I elaborate on why the talking head format is interesting from a theoretical point of view and present the relevant research on talking heads, I will first provide a more detailed description of this format.

### 1.2. Talking Heads

In terms of this dissertation project, I defined talking heads as a format of instructor presence in which learners only see the instructor's head, with the instructor lecturing into the camera (cf. Hansch et al., 2015). Talking heads may be integrated directly into the educational video (e.g., Kizilcec et al., 2014) or embedded as a picture-in-picture presentation (e.g., Kokoç et al., 2020) at various positions in the learning material. The talking head format is often used in combination with slides (Polat, 2023), for instance in lecture videos. Whereas the talking head format generally involves potential social cues such as eye contact, mimics, and lip movements (e.g., Võ et al., 2012), it does not include the instructor’s arms and therefore does not involve gestures. It is almost surprising that the format was little researched despite its frequent use (for example in lecture videos, Chen & Wu, 2015; Kizilcec et al., 2015), because it brings together an interesting combination of features: Whereas learners in the talking head format see the instructor's face, which potentially attracts visual attention and can act as a social cue, many potential advantages of instructor presence (e.g., referential cues through gestures, demonstration of actions) do not play a role in this format. As a result, talking heads offer an interesting opportunity to study the social aspects of instructor presence in relative isolation

[1] It should be noted that in certain studies (e.g., Homer et al., 2008; Moreno et al., 2001, Experiment 5) the specific format of instructor presence could not be definitively determined since the respective works did not include sample screenshots of the materials.

from other potentially learning-promoting effects. This interesting combination of features in the talking head format described above also becomes apparent when the underlying theories are taken into account.

### 1.3. Theoretical Underpinnings

When it comes to the design of instructional material, different cognitive theories are of central importance. These include the cognitive load theory (CLT) developed by Sweller and colleagues to provide guidance on how information should be presented to optimise learning, referred to in the theory as schema construction (e.g., Sweller, 1999; Sweller et al., 2011; Sweller & Chandler, 1994; Sweller et al., 1998). The theory assumes that the capacity of working memory is limited (Baddeley, 1992), whereby it can only hold about seven elements of information simultaneously (Miller, 1956). Working memory processes such as organising or comparing further reduce the number of concurrent information elements, as do interactions between the individual elements. Hence, when designing learning materials, it is important to reduce working memory load, since according to the CLT, any material design that does not consider the limitations of the working memory is deficient (Sweller et al., 1998).

In contrast to working memory, the capacities of long-term memory are unlimited following the CLT. Long-term memory not only stores isolated facts, but also complex schemas which organise and categorise individual information elements and their interactions (e.g., Sweller et al., 1998). Due to their organisation, schemas reduce working memory load as multiple elements organised into one schema can be handled as a single element in working memory, with sub-schemas no longer requiring additional working memory capacity (e.g., Sweller et al., 1998; Sweller & Chandler, 1994). Thus, the construction of schemas is one primary goal of instruction according to CLT. Although schemas are stored in long-term memory, the working memory has to process information to construct these schemas. Therefore, facilitating the processing of information in working memory and reducing working memory load is a main objective of CLT (Sweller et al., 1998).

Following CLT, working memory load depends on different factors: On the one hand, the materials themselves, i.e., their intrinsic nature, cause load, referred to as intrinsic cognitive load. One factor that influences the intrinsic cognitive load is the interactivity of the elements, i.e., whether learners can learn different elements in isolation from each other (low interactivity) or not (high interactivity), with high interactivity resulting in high intrinsic cognitive load (e.g., Sweller et al., 1998; Sweller & Chandler, 1994). As a second factor, the learners' prior

knowledge on the respective topic affects intrinsic cognitive load, with higher prior knowledge leading to lower intrinsic cognitive load (Sweller et al., 1998).

On the other hand, the way in which materials are presented also determines cognitive load, defined as extraneous cognitive load. In contrast to intrinsic cognitive load, CLT considers extraneous cognitive load as unnecessary load arising from the instructional design. More specifically, extraneous cognitive load describes the effort required to process the instructional design of the material. Consequently, according to CLT, instructional interventions can influence extraneous cognitive load, but not intrinsic cognitive load (Sweller et al., 1998), whereby intrinsic and extraneous cognitive load are additive. If extraneous cognitive load is high due to suboptimal design, but intrinsic cognitive load is rather low due to low interactivity or high prior knowledge, this instructional design does not necessarily have negative consequences for learning. However, if an instructional material poses both high extraneous and intrinsic cognitive load, it can be detrimental for learning since both sources of cognitive load taken together may exceed working memory capacity (e.g., Sweller et al., 1998).

The CLT describes germane cognitive load as the third important type of cognitive load, which supports learning by contributing to the construction of schemas (e.g., Sweller et al., 1998). Instructional design can be improved by encouraging learners to engage in conscious cognitive processing, which can be achieved by increasing learners' motivation to learn the material and making sense of the material (e.g., Mayer, 2022), for instance by incorporating social cues. However, enhancing germane cognitive load is only beneficial if the overall load remains within the working memory limits. Whereas the original version of CLT (e.g., Sweller et al., 1998) considered germane cognitive load to be a third source of cognitive load contributing to the total cognitive load, Sweller et al. (2019), in the course of further developing their theory, now assume a redistribution of resources rather than germane cognitive load imposing load itself. This involves a redistribution of working memory resources from extraneous activities to activities that are directly relevant to learning by engaging with information intrinsic to the task (Sweller et al., 2019).

The CLT is closely related to the cognitive theory of multimedia learning (CTML, e.g., Mayer, 2021), which also provides a framework for instructional design. CTML was developed by Richard Mayer, who continuously expanded the theoretical base and research base since then (e.g., Mayer, 2021). Mayer and Fiorella (2022) refer to multimedia learning as learning from words (in spoken or written form) and pictures (both static and animated pictures, graphics, diagrams, or images). According to Mayer and Fiorella (2022), meaningful learning

occurs when instruction is adapted to the functioning of the human mind. Closely intertwined with the CLT (e.g., Sweller et al., 1998), the aims of CTML include reducing extraneous processing, managing essential processing, and fostering generative learning (Mayer, 2022).

The CTML bases its framework on three underlying assumptions from cognitive science: dual channels, limited capacity, and active processing. First, based on dual-coding theory by Paivio (1990) and strongly linked to Baddeley's model of working memory (1992), CTML assumes that learners possess two different channels through which they process visual and auditory material separately (Mayer & Moreno, 2003). Learners process information such as illustrations or on-screen text in the visual channel, whereas they process spoken narrations or sounds in the auditory channel. Second, CTML assumes that the amount of information that each channel can process at one time is limited (cf. Baddeley, 1992; Sweller et al., 2011). Third, CTML considers learning as an active process that involves the selection of information, its organization into mental representations, and its integration with prior knowledge (Mayer & Fiorella, 2022).

Based on these assumptions, Mayer and colleagues developed various principles of multimedia learning that are intended to provide assistance in the design of multimedia instruction. Certain principles (e.g., image principle, coherence principle) - as well as additions to CTML (i.e., social agency theory, Mayer & DaPra, 2012; Mayer et al., 2004) - are of direct relevance to the question of whether to include talking heads in educational videos. In the following, I will first present principles that neither suggest clear advantages nor disadvantages of talking heads (i.e., image and embodiment principle), followed by those that point to negative effects (i.e., coherence and split-attention principle), before I will conclude by discussing theories and approaches that claim positive effects (i.e., social agency theory and social-cue hypothesis). The presentation of the respective principles and theories is always followed by a presentation of previous talking head research supporting the respective principles or theories.

### *1.3.1. Principles Suggesting Neither Advantages nor Disadvantages of Talking Heads*

#### 1.3.1.1. Image Principle

One principle of CTML that is important when it comes to the inclusion of talking heads in educational videos is the image principle. It states that "people do not learn better from multimedia presentations when a static image of the instructor is added to the screen" (Mayer, 2021, p. 331). It should be noted, however, that in his explanations of the image principle and the studies he uses to substantiate the image principle, Mayer (2021) refers not only to a static image of the instructor, but also explicitly to "a video of a talking head" (p. 335) as a rather

static representation of an instructor, without potential advantages such as human-like movements and gestures. The rationale behind this principle suggests that the instructor may be distracting and is not effective due to the absence of features such as gestures (Mayer, 2021). Nevertheless, it can be argued that the face itself can also convey important social cues through, for example, mimics and lip movements (e.g., Võ et al., 2012), which is not explicitly discussed in the rationale of the image principle. As support for this principle, Mayer (2021) lists seven studies which, taken together, do not provide strong evidence for adding a speaker's image on the screen. However, the majority of the comparisons refer to studies investigating animated pedagogical agents with only two of the seven comparisons concerning human instructors. In addition, many of these studies were conducted 10 to 20 years ago, whereas the technology and the production of videos has developed enormously in recent years. However, reviewing current literature on talking heads revealed a similar pattern of findings, as talking head effects remained largely absent (Kizilcec et al., 2014; Ng & Przybylek, 2021; Rosenthal & Walker, 2020). When describing the boundary conditions of the image principle, Mayer (2021) refers to the embodiment principle by specifying that certain instructors might actually be effective.

#### 1.3.1.2. Embodiment Principle

According to Mayer (2021), embodiment concerns "the ways that on-screen instructors can use their bodies to enhance the act of instructional communication" (p. 344). When an instructor displays high embodiment (e.g., by showing gestures, drawing contents), students should learn more deeply compared to when the instructor displays low embodiment (e.g., by standing still), following the embodiment principle (Mayer, 2021). Fiorella (2022) explains this learning advantage by the fact that with a high level of embodiment, "new concepts are explicitly linked to relevant actions such as hand gestures or object manipulations" (p. 286). This does not necessarily require students to enact movements themselves; they can also observe them from an on-screen instructor (Fiorella, 2022). Related to this, Fiorella and Mayer (2022a) also assume that the "effectiveness of pedagogical agents may depend on the extent to which they exhibit human-like social cues, such as gestures and facial expressions" (p. 278 f.). The rationale for this principle is that a high-embodied instructor may be more likely to serve as a social cue and thus promote deeper processing (Mayer, 2021). A high embodiment can for instance manifest in an instructor showing hand gestures instead of standing still, maintaining eye contact instead of averting gaze, or drawing contents instead of presenting contents already drawn (Mayer, 2021). In this respect, it should be emphasised that embodiment cannot easily be divided dichotomously into either high or low, but that different gradations must be

considered in relation to the instructor's different characteristics, for instance gestures or eye contact (Beege et al., 2023).

Against this background, a talking head can be described as rather low-embodied since, apart from eye contact (cf. Kuang et al., 2023) and mimics or lip movements (cf. Võ et al., 2012), this format offers few opportunities for features enabled by high embodiment (e.g., showing gestures, drawing). In this regard, Mayer (e.g., 2021) describes the instructor's embodiment as a boundary condition for the image principle. He argues that low-embodied instructors do not offer any added value according to the image principle, whereas high-embodied instructors showing human-like movements or gestures may be effective (Mayer, 2021).

#### 1.3.1.3. Underpinning Empirical Evidence

In line with the image principle, Kizilcec et al. (2014) found no effects of a visible talking head integrated next to lecture slides for short- and medium-term recall ability in a laboratory experiment. The control condition included corresponding video segments with lecture slides and spoken narration, but without talking head. In a 10-week field experiment in a MOOC environment, Kizilcec et al. (2015, Study 2) further investigated whether the permanent visibility of a talking head in the video had different effects than a talking head only being visible for certain content (cf. strategic presentation when focus on lecture slides was not necessary), but no learning differences emerged in this experiment either. Similarly, in a laboratory experiment in which Rosenthal and Walker (2020) compared different lecture formats, they found no differences between a talking head appearing in a picture-in-picture format and a voiceover condition in terms of (short-term) knowledge retention. Correspondingly, in an online experiment by Ng and Przybylek (2021), the talking head in a picture-in-picture format next to narrated slides offered no advantages over an identical video without talking head in terms of learning outcomes (recall performance). Hew and Lo (2020) also compared different video formats and found no learning differences between a format in which a talking head appeared alongside lecture slides and one in which a talking head appeared alongside free handwriting. However, the corresponding study (Study 2a), similar to the study by Kizilcec et al. (2015, Study 2), included no control condition without a talking head. However, a control condition is essential as a baseline to assess whether a talking head is beneficial or detrimental compared to an identical representation without a talking head.

### *1.3.2. Principles Suggesting Disadvantages of Talking Heads*

#### 1.3.2.1. Coherence Principle

In explaining the theoretical rationale for the image principle, Mayer (2021) refers to the coherence principle (e.g., Fiorella & Mayer, 2022b; Mayer, 2021) of CTML, which can be violated by a visible instructor. It claims that "people learn better when extraneous material is excluded rather than included" (Mayer, 2021, p. 143). The authors define extraneous material as potentially interesting but irrelevant or unneeded elements (e.g., words, images, symbols, sounds) that do not contribute to the task itself (Mayer, 2021; Mayer & Fiorella, 2014). They base their principle on the fact that such irrelevant elements compete for cognitive resources, divert attention, and disrupt learners in organising the materials. Mayer (2021) provides support for this principle with numerous studies in which concise materials (i.e., extraneous material excluded) had beneficial effects on learning outcomes. Strongly intertwined with the coherence principle, there is a lot of research on so-called seductive details, defined as interesting but irrelevant details (Garner et al., 1989). In line with the coherence principle, such seductive details (e.g., decorative images), which may distract learners' attention and enhance extraneous cognitive load, had detrimental effects on learning in numerous studies (e.g., Rey, 2012; Sundararajan & Adesope, 2020).

It could be argued that talking heads, which - as defined above - do not provide didactic functions such as gesturing or writing, do not directly contribute to the task since learners would receive all important information even without their visibility. In addition, corresponding material could also be realised without a visible talking head by simply presenting the video including the spoken narration without a visible instructor (representing a format that often constitutes the control condition compared to the instructor presence condition in research). It is further reasonable to assume that a talking head – like a seductive detail - represents an interesting element for learners since faces usually attract visual attention (Gullberg & Holmqvist, 2006; Johnson et al., 1991; Võ et al., 2012). Conversely, however, one could also assume that the talking head does not represent irrelevant or extraneous material but rather provides useful information due to its social function, giving relevance to the split-attention principle.

#### 1.3.2.2. Split-Attention Principle

In addition, the split-attention principle of CTML (Ayres & Sweller, 2022) also suggests potential disadvantages of a visible talking head in educational videos. According to this principle, designers of educational videos should avoid materials that include multiple sources

of information between which learners have to split their attention. The authors justify this recommendation by arguing that the learners' attempt to mentally integrate multiple sources of information generates extraneous cognitive load, strains the capacity of the working memory, and hinders learning (Ayres & Sweller, 2022). Ayres and Sweller (2022) report a broad body of evidence supporting the split-attention principle, also including two meta-analyses (Ginns, 2006; Schroeder & Cenkci, 2018) showing that the reduction of split-attention is associated with learning gains. However, the split-attention principle explicitly refers to materials "where each source of information is essential for understanding the material" (Ayres & Sweller, 2022, p. 199). Since the talking head hardly provides any information relevant to learning, but merely functions as a social cue, it is questionable whether the talking head can be considered a relevant source of information and thus cause split-attention in the narrow sense of the split-attention principle.

#### 1.3.2.3. Underpinning Empirical Evidence

Existing research does not enable any clear conclusions in this regard. In terms of cognitive load measures, very few studies investigated the influence of talking heads (Kizilcec et al., 2015; Ng & Przybylek, 2021). Whereas Ng and Przybylek (2021) observed no effects of a talking head on extraneous cognitive load, learners in a MOOC environment stated that they had to exert less effort when watching videos with a talking head after freely selecting the video format beforehand (Kizilcec et al., 2015, Study 1). With regard to instructor presence research in general (also including higher-embodied instructors), there were also load reducing effects of a visible instructor (Wang, Antonenko, Keil, & Dawson, 2020). In line with the coherence principle, two studies (Homer et al., 2008; Hong et al., 2018) reported higher ratings of task difficulty or mental effort (as indicators for cognitive load) due a visible instructor, although the description of the materials in both studies does not allow any clear conclusions about the examined format of instructor presence.

Previous talking head research also revealed inconsistent findings in terms of learning outcomes, which should tend to decrease due to a visible talking head according to coherence and split-attention principle (Ayres & Sweller, 2022; Mayer, 2021). To the best of my knowledge, there is only one study revealing detrimental effects of a talking head (picture-in-picture) on learning outcomes in line with both principles. In their online experiment with a Chinese sample, Yuan et al. (2021) compared four different video formats with narrated slides (original talking head, face-beautified talking head, virtual talking head, and no talking head).

Learners from the no talking head condition performed better in the retention and transfer test than those in conditions with a visible talking head.

In addition to learning measures and self-reports, research on instructor presence and specifically talking heads also relies on process measures, with eye tracking playing a central role. An eye tracker records learners' eye movements while they watch videos. The corresponding eye tracking research builds on assumptions such as the immediacy assumption and the eye-mind assumption (Just & Carpenter, 1980). First, the *immediacy assumption* presumes that individuals attempt to interpret content as soon as they fixate it, implying that processing begins with fixation. Second, according to the *eye-mind assumption*, information is processed as long as it is fixated, implying that the fixation duration corresponds to the processing duration (Just & Carpenter, 1976, 1980). These assumptions serve to interpret eye movement data in light of attentional processes. Against the background of the immediacy assumption and the eye-mind assumption, the fixation duration of a content constitutes a central measure in eye tracking research. However, it should be noted that these assumptions and therefore also the resulting conclusions regarding visual attention based on eye tracking findings involve limitations. In particular, research already demonstrated that visual attention does not always correspond with individuals' gaze measures. On the one hand, fixating a certain content does not always necessarily imply processing it, because longer fixations may, for instance, also be associated with mind-wandering (e.g., Krasich et al., 2018). On the other hand, for instance research on peripheral vision showed that individuals are able to perceive information outside the fovea (e.g., Rosenholtz, 2016).

Although eye movements and fixations do not always equate to visual attention, this process measure can provide important information on the processing of videos with and without a talking head. Further, it may be particularly useful to combine this process measure with other measures such as learning measures and self-reports and relate them to each other. In the context of this thesis, visual attention is therefore treated as equivalent to the corresponding eye movements, even though it should have become clear that the recording of eye movements represents a process measure that only allows statements about visual attention with the addition of the above-mentioned assumptions.

Empirical research, clearly suggested that faces in general (Gullberg & Holmqvist, 2006; Johnson et al., 1991; Võ et al., 2012), and talking heads in educational videos in particular (Kizilcec et al., 2014; Zhang & Yang, 2022), attracted visual attention with participants spending considerable time fixating on faces in relevant studies. Similar findings were also

evident for other types of instructors in educational videos (Colliot & Jamet, 2018; Pi & Hong, 2016; van Wermeskerken et al., 2018; Wang & Antonenko, 2017; Wang, Antonenko, & Dawson, 2020). At the same time, the fact that learners spent considerable time fixating the faces was usually associated with learners spending less time fixating the remaining visual learning materials beyond the instructor (Colliot & Jamet, 2018; van Wermeskerken et al., 2018; Wang, Antonenko, & Dawson, 2020; Zhang & Yang, 2022). A recent meta-analysis on instructor's gaze also revealed similar findings by showing that an instructor's direct gaze into the camera - which is also often evident in the talking head format - can lead to lower fixation time of the remaining learning materials (Kuang et al., 2023), which is also supported by research on gaze guidance (e.g., Pi et al., 2020; Polat, 2023; Stull et al., 2021; Wang et al., 2019). In addition, the findings of Võ et al. (2012) suggested that not only the eyes but also an instructor's mouth (i.e., lip movements) may attract attention. They found that fixations on a speaking mouth increased when participants heard an audio speech track, while fixations on the mouth and on the face in general decreased when only background music was heard instead of the audio speech track. In addition, the findings of van Wermeskerken et al. (2018) suggested that the time spent fixating an instructor hardly decreased with increasing video duration. Further, Zhang et al. (2020) showed that increasing allocation of fixations to an instructor was associated with mind-wandering, thus supporting the rationale of distraction by the talking head.

However, some questions remain unanswered with regard to eye tracking research, especially in terms of talking heads. For instance, the findings of Zhang and Yang (2022) demonstrated that a talking head only resulted in a reduction in visual attention (i.e., fixation count, dwell time) for text-based visual learning content, whereas they observed no distracting effects for graphic-based learning content (i.e., pictures). However, in their study, text-based and graphic-based learning contents were presented on the same slide. Hence, previous research left open the question of whether a talking head presented next to exclusively graphic-based visual learning content also results in reduced visual attention on the content.

In addition, van Wermeskerken et al. (2018) investigated the extent to which an instructor influenced fixations of the learning content appearing on a slide step by step. Content in educational videos often appears stepwise (i.e., sequentially) because motion attracts attention (Abrams & Christ, 2003) and thus the sequential fade-in may draw learners' visual attention to the content addressed in the spoken narration, counteracting the instructor's potentially distracting effect. Van Wermeskerken et al. (2018) found that learners fixated on

the content discussed in the spoken narration for a shorter period of time and less frequently due to a visible instructor (upper body), whereas the instructor did not affect the time to first fixation of this content. Besides fixation duration as the central gaze measure, van Wermeskerken et al. (2018) included the number of fixations as another general measure (Holmqvist & Andersson, 2017), which reflects the "fixation density" (Henderson et al., 1999, p. 211). Further, they recorded the time to first fixation defined as a latency measure, with a short time to first fixation reflecting "higher efficiency in locating the stimulus in question" (Holmqvist & Andersson, 2017, p. 591). To the best of my knowledge, no corresponding investigation using the mentioned measures exists for the talking head format. In addition, a few studies on instructor presence (van Wermeskerken et al., 2018; Wang, Antonenko, & Dawson, 2020) already investigated whether learners' eye movements (e.g., fixation time, time to first fixation) were related to their learning outcomes and found no corresponding evidence. However, to the best of my knowledge, such a study is still lacking for the talking head format, where an increasing focus on the instructor may have comparatively few advantages, as a talking head, for instance, does not provide any gestures that may guide learners' attention.

To sum up, the split-attention principle (Ayres & Sweller, 2022) may not be transferable in its original form to talking heads in educational videos since it is unclear whether a talking head constitutes an essential source of information. Nonetheless, empirical research (e.g., Kizilcec et al., 2014) already suggested that learners evidently split their attention between the visible instructor and the visual learning material next to the instructor when watching corresponding educational videos. Since the instructor does not necessarily serve as a visual source of information, this behaviour could potentially be even more detrimental to learning than a split-attention effect in its original form, because it could not only result in increased extraneous cognitive load but also in distraction from important visual learning content by the talking head. In this regard, Wilson et al. (2018) assumed that an instructor's distracting effect may be particularly detrimental next to more visual lectures than next to rather text-based contents, which are also covered by the spoken narration. However, they did not test this assumption.

Taken together, both the coherence (e.g., Mayer, 2021) and the split-attention principle (Ayres & Sweller, 2022) raise the question to what extent the talking head represents a relevant part of the learning material. If one considers the talking head to be rather irrelevant, the coherence principle suggests the exclusion of irrelevant material. If one considers the talking head to be a relevant source of (visual) information (e.g., due to lip movements), the split-attention principle suggests designing the materials in such a way that learners are not required

to split their attention. Both principles suggest similar effects with the talking head increasing extraneous cognitive load and decreasing learning outcomes, which are only weakly reflected in previous research. Whereas coherence and split-attention principle point to the disadvantages of a talking head, potential advantages are also conceivable against the background of theory.

### *1.3.3. Theories and Approaches Suggesting Advantages of Talking Heads*

#### 1.3.3.1. Social Agency Theory

While multimedia research focused for a long time on the reduction of extraneous processing and the management of essential processing in order to optimise cognitive processing, motivational and affective factors are becoming increasingly relevant as a pathway to foster generative learning (Mayer, 2022). For this reason, multimedia learning increasingly aimed to incorporate motivational and affective factors to provide a more extensive picture of how multimedia learning works. Motivation may be influenced, for instance, by affect, interest, or feelings of social connection and becomes apparent when learners engage in cognitive processing (Mayer, 2022). Affect concerns “the learner’s emotional state during learning” (Mayer, 2022, p. 70) and Moreno (2007) emphasised the importance of this affective component in the context of the cognitive affective theory of learning with media. With regard to social processes, different social cues (instructor's speech, presence, body movements) can play a role. Such social cues are supposed to generate a feeling of social connection or partnership with the instructor (Mayer, 2022; Reeves & Nass, 1996). They also play an important role because many aspects of learning with multimedia (including learning with videos) are associated with isolation or impersonality (Fiorella & Mayer, 2022a).

Efforts to consider social processes in multimedia learning build on social agency theory (Mayer & DaPra, 2012; Mayer et al., 2004), which describes how social cues can influence multimedia learning: Social cues in the learning material may trigger social responses (e.g., paying attention in order to be polite), which may lead to an increase in active cognitive processing and higher transfer performance (Fiorella & Mayer, 2022a). Fiorella and Mayer (2022a) describe social responses as a key component of social agency theory, in which learners have a feeling of social presence, which is defined as the "degree of salience of the other person in the interaction" (Short et al., 1976, p. 65). As a result, they feel connected to the instructor and invest more effort to listen to the instructional message. Since learners work harder to select, organise, and integrate information, social cues should result in better learning outcomes. Social agency theory is also taken up by the cognitive-affective-social theory of learning in digital environments (Schneider et al., 2022), which postulates similar mechanisms by

assuming that social cues in digital learning materials trigger social schemas in learners that lead to improved (para-)social, motivational, emotional, and metacognitive processes.

#### 1.3.3.2. Social-Cue Hypothesis

Besides potential effects on learning measures, instructor presence research also considers how much learners like videos with and without a visible instructor. Especially in informal learning contexts, it can be of central importance how learners experience learning with a video, as this may also influence learners' decision to select a video or continue watching it (Guo et al., 2014). In this context, the social-cue hypothesis (Colliot & Jamet, 2018) assumes that social cues in the learning material may not only increase social presence, but also contribute to a better learning experience (e.g., engagement, motivation, interest).

#### 1.3.3.3. Underpinning Empirical Evidence

Supporting social agency theory (Mayer & DaPra, 2012), Kokoç et al. (2020) observed a learning-promoting effect of a talking head in a laboratory experiment. They investigated a talking head embedded by a picture-in-picture presentation into narrated slides. In addition, a study by Zhang and Yang (2022) with a Chinese sample revealed beneficial effects of a talking head on learning outcomes. They compared three formats in a laboratory experiment and found that a talking head in combination with a narrated slideshow resulted in better comprehension test performance than a narrated slideshow alone or a narrated slideshow only including the instructor's picture.

Whereas the evidence for learning-enhancing effects of a talking head is rather limited, research on instructor presence in general (also including higher-embodied instructors) revealed a more positive pattern. Besides single detrimental effects (Wilson et al., 2018) and some absent findings (e.g., Homer et al., 2008; van Wermeskerken et al., 2018), previous research also revealed numerous positive effects of a visible instructor on learning measures (e.g., Colliot & Jamet, 2018; Hong et al. 2018; Pi & Hong, 2016; Wang & Antonenko, 2017; Wang, Antonenko, & Dawson, 2020; Wang, Antonenko, Keil, & Dawson, 2020). For example, the instructor positively affected transfer performance in general (Wang, Antonenko, Keil, & Dawson, 2020), transfer performance for difficult video topics (Wang, Antonenko, & Dawson, 2020), retention of spoken explanations (Colliot & Jamet, 2018), or recall performance for easy video topics (Wang & Antonenko, 2017). The fact that these learning-promoting effects occur particularly frequently in instructor presence research in general (including higher-embodied instructors) provides support for the embodiment principle of CTML (e.g., Fiorella, 2022). Additionally, some of these findings (Colliot & Jamet, 2018; Wang & Antonenko, 2017; Wang, Antonenko,

& Dawson, 2020) suggest that characteristics of the learning material may moderate instructor presence effects.

The extent to which a visible instructor as a social cue may increase social presence, as postulated by the social agency theory (e.g., Mayer & DaPra, 2012), remained mostly unclear in previous research. Previous investigations which included social presence measures mostly revealed no effects of a visible instructor on social presence ratings (Colliot & Jamet, 2018; Homer et al., 2008), which is also reflected by talking head research (Ng & Przybylek, 2021; Rosenthal & Walker, 2020, Study 1; Yuan et al., 2021). However, in one study (Rosenthal & Walker, 2020, Study 2), a talking head enhanced ratings of social presence compared to a voiceover condition without visible talking head.

Existing talking head research which included dimensions beyond learning (e.g., affective ratings) largely supported the social-cue hypothesis, albeit empirical evidence concerning the talking head format remained limited. In the MOOC environment investigated by Kizilcec et al. (2015, Study 1), learners stated that they had a better experience (i.e., higher affective measures) and learned more with videos including a talking head. However, the assignment to videos with and without instructors was not randomised here, but the learners selected the formats themselves. In the study by Kizilcec et al. (2014), learners also preferred formats with a talking head over those without a talking head and rated them as more helpful and useful in an affective rating. In addition, the study by Yuan et al. (2021) suggested higher satisfaction ratings for videos including a talking head, at least for one subscale, whereas the study by Rosenthal and Walker (2020) observed no talking head effects on positive emotions.

If research on instructor presence in general is also taken into account, mostly positive effects of a visible instructor on various evaluation measures become apparent (Alemdag, 2022; Henderson & Schroeder, 2021; Polat, 2023). For instance, a visible instructor positively affected ratings of satisfaction (e.g., Wang & Antonenko, 2017; Wang, Antonenko, & Dawson, 2020), perceived learning (e.g., Wang & Antonenko, 2017; Wilson et al., 2018), and situational interest (Wang, Antonenko, & Dawson, 2020; Wilson et al., 2018). Additionally, initial evidence suggested that learners were more likely to select videos with a visible instructor (Wilson et al., 2018) or talking head in a MOOC environment (Kizilcec et al., 2015) when they had a choice between different formats.

### *1.3.4. Summary of Talking Head Effects*

Taken together, regarding learning outcomes, most of the reported studies (Kizilcec et al., 2014; Ng & Przybylek, 2021; Rosenthal & Walker, 2020) showed no effects of talking

heads on learning outcomes such as retention or short- and medium-term recall (Kizilcec et al., 2014), two (Kokoç et al., 2020; Zhang & Yang, 2022) revealed learning-promoting effects (e.g., on comprehension, Zhang & Yang, 2022), and only one (Yuan et al., 2021) showed detrimental effects on learning outcomes, i.e., recall and transfer performance. This is of particular interest because, in terms of eye movement measures, previous research suggested that the talking head attracted attention (Kizilcec et al., 2014) and reduced visual attention on the learning content (Zhang & Yang, 2022). In terms of cognitive load measures, research is still very limited, with one study showing no effects (Ng & Przybylek, 2021) and one suggesting that learners need to invest less effort due to a talking head (Kizilcec et al., 2015). Similarly, no consistent pattern emerged in relation to social presence ratings, with one study showing an enhancing effect of the talking head (Rosenthal & Walker, 2020, Study 2), but in other cases the talking head had no effect (Ng & Przybylek, 2021; Rosenthal & Walker, 2020, Study 1; Yuan et al., 2021). With regard to other subjective ratings, previous talking head research reported some positive effects on affective measures (Kizilcec et al., 2014; Kizilcec et al., 2015, Yuan et al., 2021), but also absent effects (Rosenthal & Walker, 2020). The heterogeneous research evidence for talking heads suggests that certain factors may influence talking head effects.

### 1.4. Potential Factors Influencing Talking Head Effects

Various studies reported that the material next to the instructor may play a moderating role (e.g., Colliot & Jamet, 2018; Wang & Antonenko, 2017; Wang, Antonenko, & Dawson, 2020). In this regard, Wilson et al. (2018) already considered, without testing this assumption, that an instructor's distracting effect may be particularly detrimental next to visual contents, whereas an instructor next to more text-based contents, which are also covered by the spoken narration, should not be detrimental. It is conceivable that for certain learning materials it was possibly not even necessary to look at the visual content in order to learn successfully. So far, to the best of my knowledge, no study systematically investigated the extent to which the visual learning materials alongside a talking head were relevant to learning. That is, whether it was necessary to look at the learning content in order to learn successfully or whether the spoken narration was sufficient for successful learning.

Since talking heads seem to have a strong influence on learners' visual attention (e.g., Kizilcec et al., 2014), it is also conceivable that other video design features, which are also intended to direct learners' visual attention, may influence talking head effects. In terms of slides, animations allow the content to appear on the slides gradually rather than all at once. Such a stepwise fade-in appears useful against the background of various multimedia principles

(e.g., signaling, Mayer, 2021; van Gog, 2022). According to the signaling principle (Mayer, 2021), a sequential fade-in may guide learners' attention to the content currently being discussed in the spoken narration since motion attracts attention (e.g., Abrams & Christ, 2003). A sequential presentation of content may therefore counteract the potentially distracting effect of the talking head. However, to the best of my knowledge, previous research did not investigate the extent to which this presentation type (sequential vs. static presentation) interacts with talking head effects.

In addition to characteristics of the learning materials, characteristics of the learners themselves may also explain potential talking head effects. In their analysis of eye movements, Kizilcec et al. (2014) showed that learners differed greatly in the amount of time they spent looking at the talking head, ranging between 9% and 60%. This suggests that individual gaze behaviour may also represent a factor that explains talking head effects. However, to the best of my knowledge, no study investigated whether learners' eye movements explain talking head effects on learning outcomes.

### 1.5. Objective and Research Questions

Taken together, the presentation of the theoretical background revealed that talking heads are an important format, as they may potentially increase extraneous cognitive load and distract from other visual content (e.g., Ayres & Sweller, 2022), but at the same time may increase social presence as a social cue and may thus trigger deeper processing and enhanced learning (Mayer & DaPra, 2012; Mayer et al., 2004). Unlike other instructor presence formats, talking heads have a comparatively low level of embodiment (e.g., Mayer, 2021), which also implies that learning-promoting features such as referential gestures (e.g., Polat, 2023) do not play a role in talking heads. This allows the talking head format to provide an opportunity to examine the instructor's impact as a social cue in relative isolation. At the same time, the presentation of the empirical research basis showed that there exists very little research on the talking head format (e.g., Kizilcec et al., 2014; Rosenthal & Walker, 2020), which generally requires a broader body of research, to which this thesis aims to contribute. Consequently, the overarching research question focused on how a talking head influences learning outcomes (Research Question 1).

Additionally, previous research using eye tracking revealed that learners spent a lot of time looking at talking heads (Kizilcec et al., 2014) and instructors in general (e.g., van Wermeskerken et al., 2018), thus reducing visual attention on the remaining learning content. However, the question of whether this distraction by the talking head can also hinder learning

remained largely unanswered by prior research, especially for learning materials that require learners to look at the visual content in order to learn successfully. Hence, another aim of this thesis was to determine potential boundary conditions of talking heads, by investigating whether talking head effects on learning outcomes depend on the visual slide content and whether there are different effects for graphic- and text-based contents (Research Question 2). In addition, this thesis should address another potential boundary condition of talking heads concerning the learning materials, namely whether the way in which content appears on the slides (sequential vs. static presentation), impacts talking head effects (Research Question 3).

Further, the effects of a talking head on other potentially (learning) relevant variables remained largely unclear in previous research. Talking head research to date provided very few findings with regard to both cognitive load measures (Kizilcec et al., 2015; Ng & Przybylek, 2021) and social presence measures (Ng & Przybylek, 2021; Rosenthal & Walker, 2020), which represent important variables for determining explanatory mechanisms in relation to CTML (e.g., Mayer, 2021), CLT (e.g., Sweller et al., 1998), or social agency theory (e.g., Mayer & DaPra, 2012; Mayer et al., 2004). The same applies to affective measures (e.g., satisfaction). Therefore, central research questions of this thesis include how a talking head affects cognitive load, social presence, and other ratings by the learners (Research Question 4).

Previous research similarly left open questions with regard to visual attention. Prior investigations did not clarify whether talking heads distract from graphic-based content and how they affect learners' eye movements when content appears step-by-step (cf. van Wermeskerken et al., 2018). Consequently, another research question was concerned with the extent to which learners' eye movements may explain talking heads effect on learning measures (Research Question 5).

## 2. Dissertation Overview

In light of the theoretical framework described above and the empirical findings observed so far, this thesis examines the effects of talking heads in educational videos with narrated slides. For this purpose, I conducted four experiments, which I reported in three manuscripts (see Appendices A, B, and C). An overview of all experiments is provided in Table 1. The aim of this thesis is to investigate the effects of talking heads in educational videos and whether the use of talking heads is affected by certain boundary conditions (e.g., visual slide content alongside the talking head). When deriving this question, both practical aspects and the theoretical background played a crucial role. From a practical point of view, many formats raise the question regarding the instructor's visibility, in particular lecturers face the question whether they should turn on the camera or not when preparing online lectures or lecture slides. From a theoretical point of view, the following initial situation was apparent: On the one hand, different theories such as CLT (e.g., Sweller et al., 1998) and CTML (e.g., Mayer, 2021) suggest that a visible talking head in the learning material can cause additional extraneous cognitive load, and the empirical findings showed that faces (also presented as talking heads, Kizilcec et al., 2014) attracted visual attention. In addition, certain beneficial effects of instructor presence (e.g., referential cues) do not come into play with the talking head format. On the other hand, empirical research at the time of my first experiment showed no (or almost no, see Wilson et al., 2018, for instructor presence in general) learning losses due to talking heads. Hence, I raised the research question of whether the talking head format in particular has limitations that need to be considered when using it.

In this thesis, I therefore addressed a potential explanation for why previous research did not observe any negative effects of talking heads on learning outcomes: Namely that the learning materials did not require people to look at visual content in order to learn successfully. For this reason, I systematically varied the content of the videos with regard to their visual relevance in Experiments 1a and 1b (Manuscript 1) and systematically tested the recognition of visual content from the videos in Experiment 3 (Manuscript 3). Additionally, I systematically varied the presentation type of the content in Experiments 2 (Manuscript 2) and 3 (Manuscript 3) since the appearance of the content (sequentially or statically all at once) may affect potentially distracting talking head effects.

**Table 1**

*Overview of the Experiments*

| | Manuscript 1 | | Manuscript 2 | Manuscript 3 |
|---|---|---|---|---|
| | Experiment 1a | Experiment 1b | Experiment 2 | Experiment 3 |
| Setting | Online | Online | Online | Laboratory |
| Design | 2 x 2 between-within-subjects | 2 x 2 x 2 between-within-subjects | 2 x 2 x 2 within-subjects | 1 x 2 within-subjects |
| Independent variables | Talking head (between, yes/no) | Talking head (between, yes/no) | Talking head (within, yes/no) | Talking head (within, yes/no) |
| | Slide type (within, graphic-/text-based) | Slide type (within, graphic-/text-based) | Slide type (within, graphic-/text-based) | |
| | | Presentation type (between, sequential/static) | Presentation type (within, sequential/static) | |
| Materials | 6 videos per participant | 4 videos per participant | 8 videos per participant | 8 videos per participant |
| Main dependent variables | Knowledge acquisition (multiple choice) | Knowledge acquisition (multiple choice) | Knowledge acquisition (free recall) | Knowledge acquisition (free recall) |
| | | | Video selection | Picture recognition (multiple choice) |
| | | | | Visual attention (e.g., fixations) |
| | Cognitive load (e.g., extraneous cognitive load) | Cognitive load (e.g., extraneous cognitive load) | Cognitive load (e.g., extraneous cognitive load) | Difficulty, invested mental effort |
| | Social presence | Social presence | Social presence | Social presence |
| | Satisfaction | Satisfaction | Satisfaction | Satisfaction |
| | Perceived learning | Perceived learning | Perceived learning | Perceived learning |
| Sample | $N$ = 96 (Prolific) | $N$ = 184 (Prolific) | $N$ = 112 (Prolific) | $N$ = 96 (University students) |
| Pre-registration | https://osf.io/vcgp5 | https://osf.io/r36wn | https://osf.io/sd54z/?view_only=43bab4086ffb45989b04f077a925627e | https://osf.io/rcde6/?view_only=fb6aa775e1b14d30826d2af29071bc51 |
| Data and scripts | https://osf.io/xug9c/?view_only=7b149959778c40c3b9fe51fb988f823f | https://osf.io/xug9c/?view_only=7b149959778c40c3b9fe51fb988f823f | https://osf.io/afnb8/?view_only=d1b698ba8427488ab3200d36c096fe83 | https://osf.io/4uwx6/?view_only=47a628aba0e54eafa7a0887b60af4266 |

In addition, with the different experiments included in this thesis, I aimed to obtain a broad picture of talking head effects on different dependent variables, since previous research on this format provided limited findings, especially on measures beyond learning outcomes. I followed this multi-criterial approach because many aspects were relevant when discussing the effects of talking heads. In this context, learning measures such as factual knowledge acquisition are of initial interest in order to determine whether a talking head influences the extraction of factual knowledge from the learning materials. Additionally, due to the link to CLT and CTML (e.g., Sweller et al., 1998; Mayer, 2021), various measures of cognitive load (especially extraneous and germane cognitive load) were also of central importance. Recording these cognitive load measures allowed testing whether the empirical evidence generated by the experiments supported the derivations from these theories. Similarly, the social agency theory (e.g., Mayer & DaPra, 2012; Mayer et al., 2004) suggests positive effects of social cues on social presence and also on learning measures. It remained an open question whether talking heads are capable of triggering such social processes despite their comparatively low embodiment (e.g., Mayer, 2021), leading me to include social presence as a central measurement in all experiments. Additionally, I was interested in talking head effects on affective variables (e.g., satisfaction) and further subjective ratings such as perceived learning, since previous research on talking heads had not yet provided broad evidence in this regard, although these variables are of particular importance in the context of informal learning. In such informal learning settings, learners often have a free choice between different learning opportunities, which may prevent them from responding to less appealing formats. Therefore, I decided to collect all these previously described measures (cognitive load, social presence, satisfaction, perceived learning) across all experiments.

I extended this multi-criterial approach in different experiments to include further variables which were of central importance for the research questions addressed by the respective experiments. I decided to use a behavioural measure by including a selection task in Experiment 2 to capture which formats learners prefer when choosing between different educational videos. Further, I aimed to include process data by recording learners' eye movements in Experiment 3 in addition to learning measures and self-reports to gain an insight into how a talking head affects actual gaze behaviour and whether these eye movements are related to learning measures. By incorporating a picture recognition task in Experiment 3, I intended to directly assess the recognition of visual learning content, which was already successfully implemented in other contexts (Glaser & Schwan, 2015).

When designing the educational videos which included the talking head, I opted for producing videos with narrated slides. Narrated slides can be described as video format in which content (e.g., graphic- or text-based) appears on slides with a spoken narration referring to it. This represents a popular format of educational videos, which is often used in online lectures. I integrated the talking head in the video slideshow by using greenscreen technology. Since many studies on instructor presence in general and talking heads in particular only included single videos or topics in their investigations, I aimed to examine the talking head with various video topics to enhance the generalisability of my findings.

To answer my research questions, I used an experimental paradigm in all studies allowing a systematic variation of the presence of a talking head. For this purpose, I produced content-equivalent videos with and without a talking head, implying that learners saw either videos consisting of slides on which a talking head was displayed to introduce the content or identical videos (identical slides, identical spoken narration) without a visible talking head on the slides. I varied the visibility of the talking head partly using a between-subjects design (Experiments 1a and 1b), with learners either seeing all videos with or all videos without a talking head, and partly using a within-subjects design (Experiments 2 and 3), with learners seeing half of the videos with a talking head and the other half without talking head, while balancing which topic was presented with and without talking head.

When selecting the experimental setting, which was partly influenced by the COVID-19 pandemic, I opted for both online experiments (Experiments 1a, 1b, and 2) and a laboratory experiment with eye tracking (Experiment 3). The online experiments are characterised by a high external validity because the learners saw the videos in a similar context in which they are usually employed. In contrast, the laboratory experiment offered higher internal validity and greater experimental control, allowing the elimination of many potential interfering variables and ensuring that occurring effects could most likely be attributed to experimental variation. The use of both online and laboratory experiments also provided the opportunity to observe whether similar talking head effects showed up in both settings despite the different contextual factors. For all experiments, I preregistered hypotheses and corresponding analyses, manipulations, design, sample sizes, measures, and exclusion criteria. All data and scripts can be downloaded at Open Science Framework (OSF). Links to all preregistrations, data, and scripts are listed in Table 1. In the following, I will describe the four experiments and their genesis. Thereby, I will only address key variables and hypotheses. For a detailed description, please refer to the corresponding manuscripts in Appendices A, B, and C.

## 2.1. Manuscript 1: Experiments 1a and 1b

### *2.1.1. Experiment 1a*

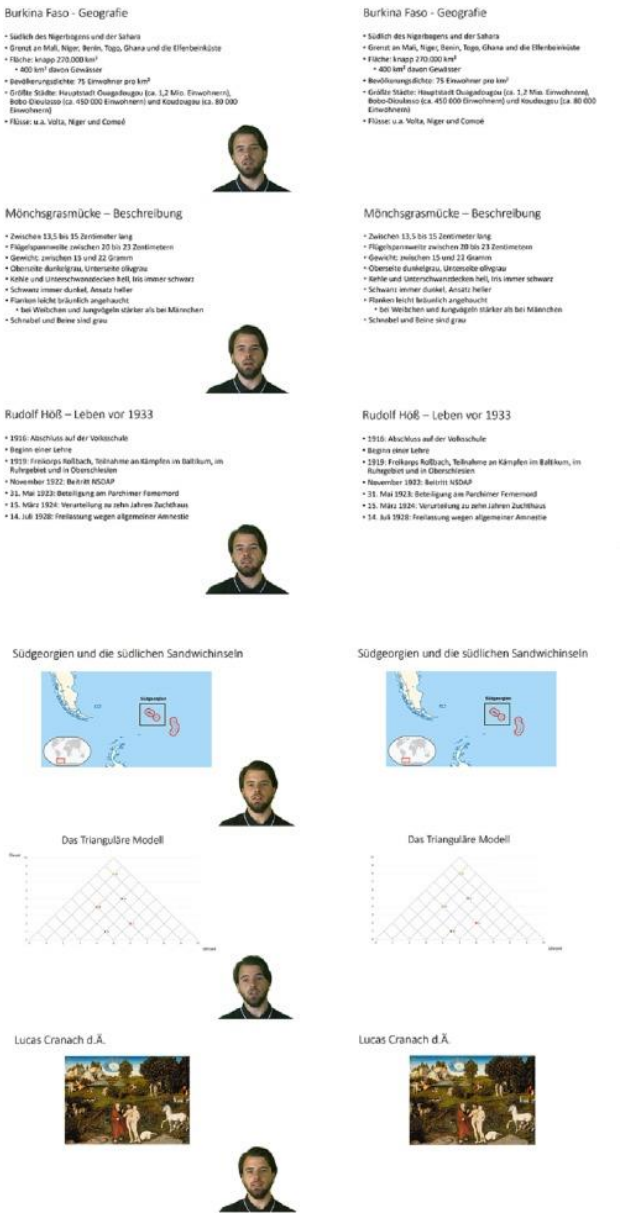
Experiment 1a was an online experiment that investigated the effects of a male talking head in six educational videos. Since talking heads and instructors in general often attracted visual attention, but rarely had learning-hindering consequences, I tested whether the talking head had different effects depending on whether the visual content was relevant (learners had to look at the content for successful learning) or not (learners did not have to look at the content for successful learning). I addressed this question using a 2x2 design, varying the presence of the talking head (present vs. absent) as a between-subjects factor and the relevance of the visual content (relevant vs. not relevant) as a within-subjects factor. In videos with talking head, the talking head appeared either in the bottom right or left corner (balanced). Operationalising the relevance of the visual content involved the use of graphic-based vs. text-based videos (on different topics). For this purpose, I produced videos on six different topics (see Figure 1). For graphic-based content (three videos), learners had to focus on the visual learning content to receive all information relevant to learning, whereas for text-based content (three videos), learners could receive all relevant information by listening to the spoken narration. In addition to learning outcomes (recall of factual information, multiple choice questions), the experiment further included cognitive load ratings (Klepsch et al., 2017), social presence ratings (adapted from Kizilcec et al., 2015; Schramm & Hartmann, 2008), and other subjective ratings such as satisfaction (adapted from Wang and Antonenko, 2017) and perceived learning (adapted from Kizilcec et al., 2015; Wang and Antonenko, 2017) as dependent variables in order to obtain a broad insight into the effects of talking heads.

I expected the talking head to result in learning losses for relevant visual content (i.e., graphic-based videos) since learners would miss important information by looking at the talking head. However, I did not expect such learning losses due to the talking head in text-based videos since learners were able to compensate for the talking head's possible visual distraction by listening to the spoken narration. Regarding subjective ratings, I expected the talking head to increase ratings of satisfaction and perceived learning based on prior research on instructor presence (e.g., Wang & Antonenko, 2017).

The study was conducted with LimeSurvey, and participants were recruited via the online panel provider Prolific. The final sample size consisted of 96 participants (26 female, $M_{age}$ = 29.01, $SD_{age}$ = 8.64). After randomised assignment and capturing learners' self-reported prior knowledge regarding the video topics, the learners watched six videos (system-paced) and

**Figure 1**

*Screenshots of videos with talking head (left) and without talking head (right) dealing with text-based (Rows 1 to 3) and graphic-based (Rows 4 to 6) topics (Experiment 1a)*




*Note*. Picture credits: All pictures except the pictures in Rows 4 and 6 were self-generated. Row 4: TUBS, South Georgia and the South Sandwich Islands in its region. This graphic contains elements which have been taken or were adapted from the following graphic (CC BY-SA 3.0): via https://commons.wikimedia.org/w/index.php?curid=15431607, Row 6: Lucas Cranach the Elder, Adam and Eve-Paradise (Public domain): via https://commons.wikimedia.org/wiki/File:Lucas_Cranach_(I)_-_Adam_and_Eve-Paradise_-_Kunsthistorisches_Museum.jpg

indicated various ratings. Due to time constraints, some ratings were recorded after each video and other ratings were recorded only once after participants had watched all videos. This was followed by the knowledge test.

Contrary to the assumption, the talking head did not affect knowledge test performance, regardless of visual content (graphic-based vs. text-based). In addition, results revealed no talking head effects on various subjective ratings (e.g., cognitive load, social presence, satisfaction, and perceived learning). I presumed that the absence of learning losses due to the talking head next to graphic-based content might be due to the content fading in sequentially (animated slides). Since motion attracts attention (Abrams & Christ, 2003), according to the signaling principle (e.g., Mayer, 2021), the sequential fade-in could have repeatedly drawn learners' attention away from the talking head and back to the visual learning content.

### *2.1.2. Experiment 1b*

To test the proposed explanation outlined in the previous section and replicate Experiment 1a, I conducted a second experiment, which additionally varied whether the content appeared on the slides sequentially step by step or statically all at once. This time, I expected the talking head to result in learning losses with static presentation of visual content, whereas I expected a replication of the effects from Experiment 1a for the content presented sequentially. Also, according to the findings from Experiment 1a, I expected no effects of the talking head regarding learners' subjective ratings.

Experiment 1b was conducted online, almost completely identical to Experiment 1a except for the additional between-subjects factor presentation type (sequential vs. static). Since it was not possible to implement the static presentation for all six video topics from Experiment 1a, the learners saw four of the six videos of the preceding experiment (two graphic-based, two text-based). The final sample consisted of 184 participants (77 female, 103 male, 4 diverse, $M_{age}$ = 28.68 years, $SD_{age}$ = 8.47).

The experiment indicated no impact of the talking head on knowledge test performance, neither for sequential presentation (replication of the findings from Experiment 1a) nor for static presentation, contrary to the hypothesis. In addition, the talking head again did not affect the majority of learners' subjective ratings (e.g., perceived learning, satisfaction, social presence), but resulted in a decrease in extraneous cognitive load. Both Experiments 1a and 1b showed hardly any impact of the talking head on most of learners' subjective ratings, although previous research often suggested an influence of talking heads (e.g., Kizilcec et al., 2014, Kizilcec et al., 2015) or instructors in general (e.g., Wang & Antonenko, 2017) in this regard. As a possible

explanation for this, I hypothesised that participants in Experiments 1a and 1b did not have a direct comparison between formats with and without talking head, unlike in other studies that observed effects of instructor presence on different ratings such as perceived learning and satisfaction (e.g., Kizilcec et al., 2014, Kizilcec et al., 2015; Wang & Antonenko, 2017). Hence, I decided to conduct a new experiment focusing more strongly on aspects of subjective ratings in a within-subjects design.

### 2.2. Manuscript 2: Experiment 2

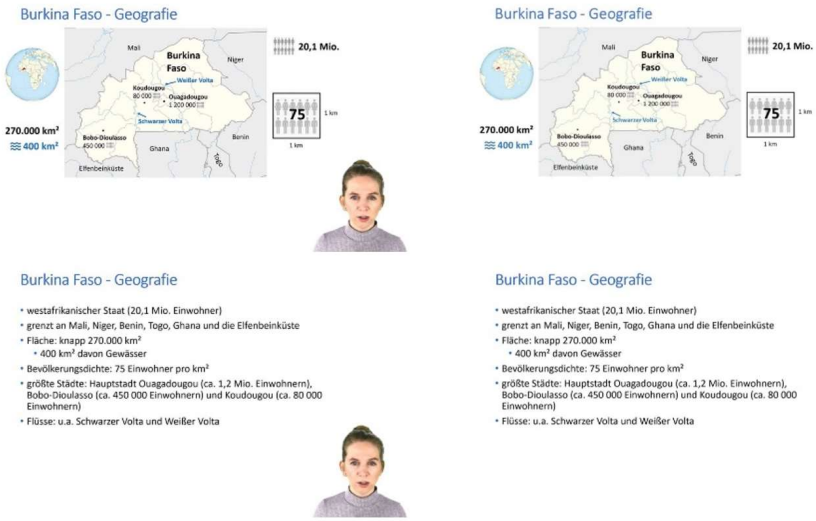
Especially in informal learning contexts, it is of central importance whether learners like a video since this may determine whether a video is selected or continued to be watched (Guo et al., 2014). Experiment 2 therefore increasingly focused on learners' subjective ratings, resulting in a variety of changes compared to the previous experiments: Experiment 2 varied the presence of a talking head as within-subjects factor to examine whether the talking head resulted in differences in subjective ratings when the learners had a direct comparison between different formats. Additionally, I implemented a selection task to determine to what extent a talking head could affect whether learners selected a video for learning. Further, I assessed learning outcomes using a free recall test of factual information instead of a multiple-choice test to potentially increase the difficulty of the knowledge test.

Whereas graphic and text-based videos dealt with different topics in Experiments 1a and 1b, I modified this for Experiment 2 by producing content-equivalent graphic and text-based videos. For this purpose, I produced new videos on eight different topics, this time including a female talking head (see Figure 2). Since the same content had to appear in both graphic- and text-based formats, slide type remained an independent variable, but the relevance of visual information as a possible moderator moved more into the background, as the previous experiments did not show any interactions with the talking head for learning outcomes, even when considering the presentation type (i.e., sequential vs. static presentation).

In addition to the slide type, I also varied the presentation type (sequential vs. static presentation), as in Experiment 1b, in order to examine the extent to which this design feature influences potential talking head effects. These three two-level factors (talking head, slide type, presentation type) were all varied as within-subjects factors, resulting in a 2x2x2 within-subjects design, with each learner seeing one video in each of the eight versions resulting from the combination of the features. For this purpose, I produced eight different video versions per topic. Between learners, I balanced which video topic they saw in which version to avoid confounds between content and version. Based on previous research in which learners had a di-

**Figure 2**

*Screenshots of videos with (left) and without (right) talking head appearing on graphic-based (top) and text-based (bottom) slides (Experiment 2)*




*Note*. Picture credits: This graphic contains elements which have been taken or were adapted from the following graphic (CC BY-SA 3.0): via https://de.wikipedia.org/wiki/Datei:Burkina_Faso_on_the_globe_(Africa_centered).svg

rect comparison between videos with and without an instructor (e.g., Wang & Antonenko, 2017), I expected positive effects of the talking head on different subjective ratings (e.g., satisfaction perceived learning). In terms of learning outcomes, based on exploratory supplemental analyses of Experiment 1b (cf. Appendix A), I expected a three-way interaction, reflecting particularly weak learning outcomes for a visible talking head with graphic-based content and static presentation.

Experiment 2 was conducted as an online experiment via LimeSurvey. The sample consisted of 112 participants recruited via Prolific (64 female, 43 male, 5 diverse, $M_{age}$ = 26.88 years, $SD_{age}$ = 9.13). After indicating their self-assessed prior knowledge, learners watched eight short videos (system-paced), with various rating measures recorded after each video. This was followed by the knowledge test, which, different from Experiments 1a and 1b, tested free recall (instead of multiple choice) of factual knowledge from the videos. Finally, there was the video selection task, before which the learners received the information that they would now watch another video and could choose the format.

According to the hypothesis, there were positive effects of the talking head on different subjective ratings such as satisfaction and perceived learning. In contrast, however, I observed

a small but negative effect of the talking head on knowledge test performance, regardless of slide type or presentation type, which was also not in line with the expected three-way-interaction. In addition, participants indicated considerably higher social presence ratings for videos including a talking head and selected formats with talking head more frequently in the selection task. Further, the talking head resulted in a decrease in extraneous cognitive load and an increase in germane cognitive load.

Whereas the aforementioned talking head effects emerged regardless of slide type and presentation type, there were consistent interactions between slide type and presentation type for several dependent variables (e.g., perceived learning, satisfaction, extraneous cognitive load), suggesting that sequential presentation was beneficial for graphic-based content, while it often had no effects for text-based content. Overall, the contrasting talking head effects imply that the decision in favour of or against a talking head might include a trade-off between educational usefulness and desired video popularity.

### 2.3.Manuscript 3: Experiment 3

In the previous online experiment, the talking head had a small but detrimental effect on learning outcomes. Whereas this online study offered high ecological validity with learners watching videos in a similar context in which they are frequently used, the experimental control was rather low, and it was not possible to record learners' eye movements. Hence, I decided to conduct a similar experiment (i.e., again varying of the talking head as within-subjects factor, identical knowledge test) in the laboratory using eye tracking to investigate whether the talking head influences learners in how long they fixated the learning content and whether their eye movements may contribute to explain learning effects. Since slide type and presentation type did not influence talking head effects on learning outcomes and other key variables in Experiment 2, I opted against systematically varying these two design features in Experiment 3.

I conducted Experiment 3 in the laboratory at the University of Tübingen. For the materials, I used a subset of the materials from Experiment 2. I decided to use graphic-based videos with sequential fade-in since the combination of visual content that appears dynamically represents an interesting combination of materials for an eye tracking study. In particular, a reason for choosing graphic-based content was the lack of empirical evidence showing a distracting effect of the talking head next to purely graphic-based learning content at the level of eye movement measures. In addition, the entirely graphic-based content allowed assessing learning outcomes in an additional format: In this experiment, I implemented a picture recognition task (cf. Glaser & Schwan, 2015) to determine whether the talking head influences

the recognition of visual details from the videos. Further, the stepwise appearing content allowed to examine whether the talking head impacts attention to newly appearing learning content, i.e., how long (fixation duration), how often (number of fixations), and how quickly (time to first fixation) learners fixated the newly appearing content on the slides (cf. van Wermeskerken et al., 2018).

I varied the presence of the talking head using a within-subjects design, with every learner watching videos on eight different topics, four of them with and four without talking head. Again, I balanced whether the talking head appeared in the bottom right or bottom left corner of the video. As this made the content appear either slightly further to the right or left in the video, I also balanced whether the content appeared further to the right or left for the conditions without a talking head. Further, it was balanced between participants which topic was presented with and without talking head.

Based on previous research using eye tracking (e.g., Kizilcec et al., 2014), I expected a decrease in content fixation duration for videos including a talking head. Following the previous online experiment using similar materials and the same knowledge test (i.e., Experiment 2), I expected better knowledge test performance for videos without talking head than for videos with talking head. Accordingly, I expected better picture recognition task performance without talking head as the talking head attracted visual attention in previous research (e.g., Kizilcec et al., 2014). Since the experiment included both learners' eye movement measures and two different learning measures (knowledge test, picture recognition task), I additionally aimed to explore whether talking head effects on learning measures were explained by content fixation duration. In addition, based on the results in Experiment 2, I expected positive talking head effects on different subjective ratings (social presence, satisfaction, perceived learning).

Ninety-six students of the University of Tübingen participated in the eye tracking experiment (73 female, $M_{age}$ = 22.61, $SD_{age}$ = 3.78). The experimental procedure included two laptops per participant, with one laptop presenting the videos during eye tracking and the other presenting the knowledge test and picture recognition task. After indicating their self-reported prior knowledge and completing the calibration procedure, learners saw eight videos (system-paced) on eight different topics while their eye movements were recorded using a Tobii Pro Nano eye tracker (60 Hz). During the complete eye tracking procedure, the participants were asked to place their head on a chinrest to improve data quality. After each video, learners answered questions regarding subjective ratings. In order to keep the eye tracking part as short as possible, there was no detailed measurement of the cognitive load after each video, but the

measurement involved only one item on perceived difficulty (as indicator for intrinsic and extraneous cognitive load) and one item on invested mental effort (as indicator for germane cognitive load). After watching all videos, learners completed the knowledge test and the picture recognition task.

In line with the hypothesis, the talking head strongly influenced learners' eye movements by reducing content fixation duration. Additionally, newly appearing learning content was fixated shorter, less frequently, and more slowly when presented next to a talking head. However, in contrast to the hypothesis, the talking head did not affect both learning measures (i.e., knowledge test and picture recognition task performance). Further, content fixation duration could not explain talking head effects on learning outcomes. Regarding subjective ratings, the talking head did not affect most of the ratings (e.g., satisfaction, perceived learning) but had a strong positive effect on ratings of social presence. Whereas there was a strong distraction by the talking head at the level of eye movements, no learning losses were observed at the level of different learning outcomes.

### 2.4. Internal Meta-Analysis

It is evident from the presentation of the experiments that the findings regarding various dependent variables (e.g., knowledge test performance, social presence, satisfaction) in the different experiments were heterogeneous, so that only limited general conclusions are possible. For this reason, additional meta-analyses were calculated for the central variables covered in all experiments in order to enable overall conclusions based on the data of this thesis. Nevertheless, it should be considered that a small number of included studies in a meta-analysis, as in this case with $N = 4$ experiments, leads to a poorer estimate and lower precision (Kelley & Kelley, 2012). A single meta-analysis was calculated for each of the central dependent variables that were recorded across all experiments, i.e. knowledge test performance, extraneous and germane cognitive load, social presence, satisfaction, and perceived learning. Please note that difficulty served as equivalent for extraneous cognitive load and invested mental effort served as equivalent for germane cognitive load in Experiment 3. When calculating the meta-analyses, I selected a random-effects model, allowing the true effect to vary from study to study, and opted against a fixed-effects model, which would have assumed one true effect size (Borenstein et al., 2009). Since not all experiments were identical in design, materials, and setting, this might have affected the respective effects. I calculated the meta-analyses using JASP (Version 0.18.3). In choosing the random-effects model, I opted for a restricted maximum likelihood method, also taking into account the degrees of freedom used in

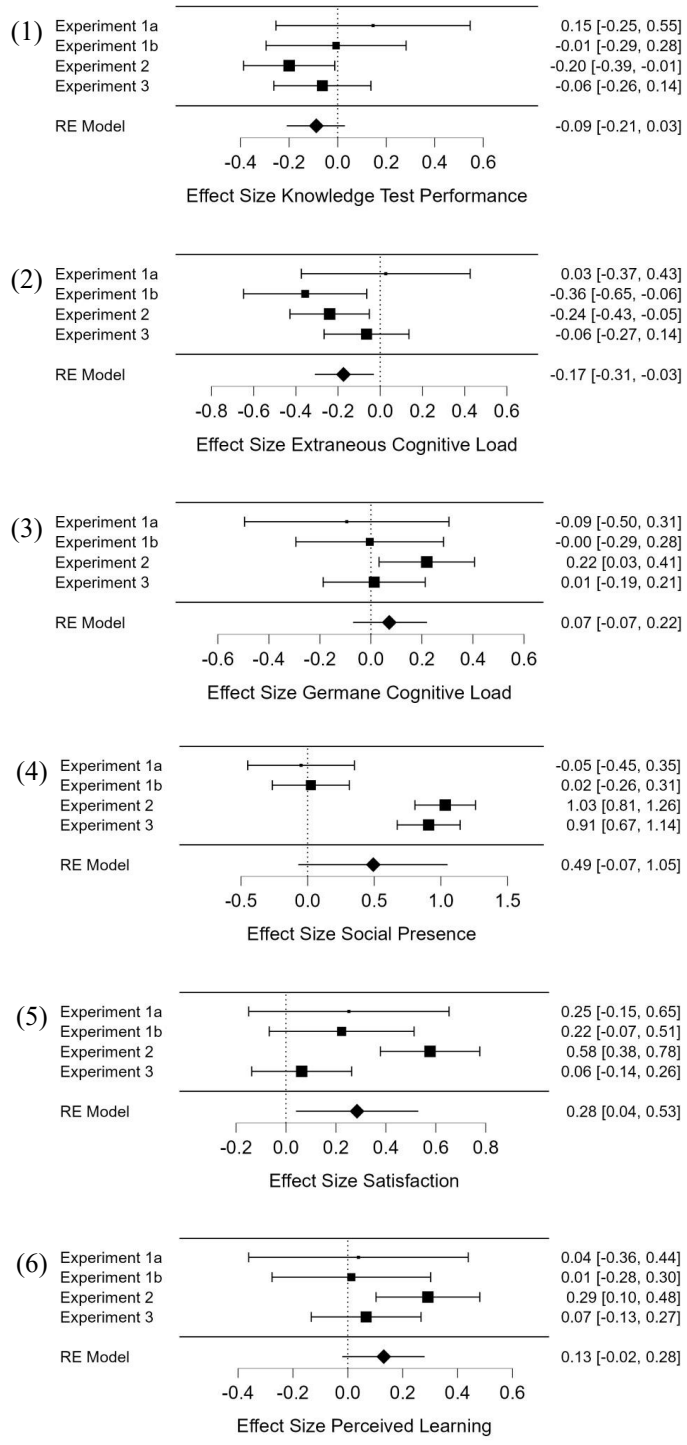
estimating the fixed-effect portion of the model, which is particularly relevant for meta-analyses including only a small numbers of studies (Kelley & Kelley, 2012). Cohen's *d* was used as the effect size, since it is considered a robust estimator for samples with $N > 20$ (Goulet-Pelletier & Cousineau, 2018), which applies to all included experiments. When interpreting effect size estimates, I refer to Cohen's (1988) classification, which categorises effects $d < 0.5$ as small, $d < 0.8$ as medium, and $d \geq 0.8$ as large effects. To obtain Cohen's *d*, SPSS was used to calculate t-tests for either independent samples (Experiments 1a and 1b) or dependent samples (Experiments 2 and 3) in order to examine only the talking head effect for the respective experiments. The standard error of the effect size *d* was calculated using the 95% confidence intervals of Cohen's *d* provided by SPSS ($\frac{\text{CI}_{\text{upper limit}} - d}{1.96}$). The corresponding JASP-file for all meta-analyses also including the relevant data can be found on OSF (https://osf.io/49xvp/?view_only=b7b18de79b64409ba839717dcc7e2a0f).

The forest plots of all conducted meta-analyses are displayed in Figure 3. For knowledge test performance, the random-effects model revealed no significant effect across the four experiments, $d$ = -0.088, $SE$ = 0.062, $p$ = .154, 95% CI [-0.209, 0.033]. However, the talking head significantly reduced ratings of extraneous cognitive load compared to formats without talking head, $d$ = -0.174, $SE$ = 0.071, $p$ = .014, 95% CI [-0.314, -0.035]. Regarding germane cognitive load, the talking head had no significant overall effect, $d$ = 0.072, $SE$ = 0.073, $p$ = .324, 95% CI [-0.071, 0.215]. Similarly, the random-effects model revealed no significant talking head effect for social presence across all four experiments, $d$ = 0.493, $SE$ = 0.285, $p$ = .083, 95% CI [-0.065, 1.050]. In contrast, a small talking head effect emerged in the meta-analysis on satisfaction, $d$ = 0.285, $SE$ = 0.125, $p$ = .022, 95% CI [0.040, 0.529], with the talking head increasing learners' satisfaction ratings compared to videos without talking head. Finally, the random-effects model on perceived learning revealed no significant effect across the four experiments, $d$ = 0.131, $SE$ = 0.077, $p$ = .090, 95% CI [-0.021, 0.283].

In summary, the talking head significantly reduced learners' ratings of extraneous cognitive load across different studies. Further, the meta-analyses showed an increase in satisfaction due to the talking head across the four experiments, while the talking head had no overall effect on the other dependent variables.

**Figure 3**

*Forst plots of the random-effects (RE) meta-analyses conducted on (1) knowledge test performance, (2) extraneous cognitive load, (3) germane cognitive load, (4) social presence, (5) satisfaction, and (6) perceived learning*



# 3. General Discussion

## 3.1. Summary and Discussion of Findings

In the following, I first summarise and discuss the central results of the experiments grouped by dependent variables. In doing so, I concentrate on the central variables and results for the overall thesis, which also constituted the focus of the introduction and the presentation of the individual experiments. For an insight into the more detailed findings, please refer to the individual manuscripts in Appendices A, B, and C.

### *3.1.1. Learning Outcomes*

With regard to Research Question 1, addressing talking head effects on learning outcomes, the reported findings on knowledge acquisition were not entirely consistent across different experiments. In the first two online experiments, varying the presence of a talking head as between-subjects factor and recording knowledge acquisition using a multiple-choice test, the talking head did not affect factual knowledge acquisition. However, in the last online experiment (Experiment 2), varying talking head as within-subjects factor and measuring knowledge acquisition using a free recall test, the talking head had a small negative effect on factual knowledge acquisition. Experiment 3, which took place in the laboratory and studied the talking head as within-subjects factor, showed no effects of the talking head using an identical free recall test as in Experiment 2. When the internal meta-analysis across the four experiments regarding knowledge test performance is also taken into account in answering Research Question 1, there was also no effect of the talking head on learning outcomes. In addition, the talking head did not affect the recognition of pictures from the videos in Experiments 3.

Another aim of this thesis was to determine whether talking head effects on learning outcomes depend on certain boundary conditions concerning the learning materials, i.e., slide type (Research Question 2) and presentation type (Research Question 3). In all experiments, talking head effects on learning outcomes were independent of other design features that were additionally varied systemically in the experiments (i.e., slide type in Experiments 1a, 1b, and 2; presentation type in Experiments 1b and 2). Hence, the findings do not provide support for the investigated boundary conditions.

Possible explanations for the different results on learning outcomes are outlined in the following. Given the similarity of the first two experiments (e.g., talking head varied as a between-subjects factor, identical talking head, similar materials, similar knowledge test), it is

hardly surprising that the first two experiments provided a very uniform pattern of findings, especially with regard to learning outcomes. Experiment 1b replicated Experiment 1a and additionally investigated a potential explanation for the absence of learning losses in Experiment 1a, namely that the sequential presentation of the content compensated for the distracting effect of the talking head, which was not supported by the results. The learning outcomes in Experiment 2 differ from those of the previous two experiments, and there are several possible reasons for this, considering the numerous differences between these experiments. For instance, the experiments included different educational videos, different talking heads, a different design with respect to the talking head (between- vs. within-subjects factor), and different knowledge tests (multiple choice vs. free recall). Finally, Experiment 3 tested a potential explanation for learning deficits in Experiment 2, namely that learners performed worse when looking at the learning content for a shorter time. However, Experiment 3 could neither replicate learning losses nor observe a mediating effect of learners' eye movements regarding talking head effects on learning measures (cf. Research Question 5).

However, in view of the numerous similarities between Experiments 2 and 3 (talking head as within-subjects factor, identical talking head, similar learning materials, identical knowledge test), the question arises as to why the learning-inhibiting effect of the talking head was not replicated. One potential explanation for this might lie in the changed experimental setting (online vs. laboratory experiment). Watching an educational video at home in familiar surroundings and without supervision represents a completely different learning situation than watching the same video in a laboratory room, with the head on a chinrest, knowing that one's eye movements are recorded and analysed. Correspondingly, the meta-analysis by Alemdag (2022) already suggested that the study setting can serve as a moderator for instructor presence effects on learning measures, with positive effects in laboratory studies, and no (but descriptively negative) effects in online studies. Hence, talking head effects appear to be sensitive to the experimental setting. Both my findings and the meta-analysis by Alemdag (2022) suggested that talking head effects tend to be more positive in the laboratory, with my experiments revealing no effect in the laboratory (Experiment 3) and a negative effect online (Experiment 2) and Alemdag (2022) showing a positive effect in the laboratory and no effect online. Hence, future research should address potential reasons for this pattern, for instance by systematically studying the participants' invested effort.

Another difference between Experiments 2 and 3 concerned the investigated sample. In Experiment 2, the sample was recruited using the online panel provider Prolific, whereas in Experiment 3, the student mailing list of the University of Tübingen served for recruitment.

Accordingly, whereas the proportion of university students in the sample of Experiment 3 was almost 100%, in Experiment 2 it was only around 54%. With regard to school education, however, there were no major differences between the experiments. In Experiment 3, all participants stated either high school diploma or university degree as their highest qualification, and in Experiment 2 this proportion was also 92%. In addition, the performance in the identical knowledge test was comparable in both experiments[2], which argues against systematic differences between the samples being responsible for the different talking head effects.

Taken together, the talking head had largely no effects on learning measures, apart from a rather small negative effect on the recall of factual information from the videos in a comparatively informal setting in Experiment 2. The internal meta-analysis also supported the predominantly absent effects on learning outcomes by revealing no overall talking head effect across the four experiments. These observations correspond to the findings in previous talking head research, where the majority of studies also showed no learning effects (Kizilcec et al., 2014; Ng & Przybylek, 2021; Rosenthal & Walker, 2020). A detrimental effect on learning measures as in Experiment 2 was only found in a talking head study by Yuan et al. (2021), which was also an online experiment with learners performing better in a retention and transfer test without a talking head. It is important to note that I did not observe any learning-promoting effects of a talking head in any of the experiments, contrary to the findings of Kokoç et al. (2020) and Zhang and Yang (2022). The absence of learning-promoting effects (even in the laboratory) is not surprising in light of the research on the embodiment principle (e.g., Fiorella, 2022), which suggests no learning advantages for low-embodied instructors such as talking heads compared to high-embodied instructors.

Further, in terms of Research Questions 2 and 3, the design features systematically varied in Experiments 1a, 1b, and 2 did not influence talking head effects on learning outcomes in any of the experiments. Instead, the talking head in Experiment 2 had a small but detrimental effect on learning, regardless of other design features. In particular, the talking head did not interfere with learning even though I systematically designed graphic slides for Experiments 1a and 1b in which crucial information relevant for the successful completion of the knowledge test was only presented visually on the slides. Correspondingly, the talking head did not affect picture recognition task performance in Experiment 3, which systematically asked for visual details that learners could only recognise by looking at the slides. In addition, eye tracking data

[2] A t-test for independent samples comparing mean scores of Experiments 2 and 3 revealed no significant difference in knowledge test performance, $t(206) = 0.162$, $p = 0.872$.

from this experiment also demonstrated that learners looked at the content for a shorter time when including a talking head. Learners therefore seemed to compensate very well for the visual distraction caused by a talking head and learned just as well even with reduced visual attention on the learning content. A potential explanation might be that even very short fixations were sufficient to successfully answer the corresponding tasks (cf. Eitel et al., 2012). Against the background of these findings, it would be interesting to examine the influence of a talking head next to learning content with a higher visual complexity (see Meier et al., 2023).

Similarly, the presentation of the content (sequential vs. static) did not systemically influence the talking head effects on learning. In this respect, the eye tracking findings from Experiment 3 also offer an interesting perspective, even though the presentation type was no longer systematically varied in Experiment 3. Nonetheless, the learners' eye movements with sequentially appearing content indicated that the talking head decreased learners' visual attention on newly appearing content in terms of various measures (i.e., fixation duration, number of fixations, time to first fixation). This at least suggested that the sequential presentation of content on the slides alone could not completely compensate for a talking head's distracting effect on the level of process data.

### *3.1.2. Visual Attention*

As eye movement measures were only recorded in one of the four experiments, no conclusions can be drawn across studies in this regard. However, consistent with previous research and a recent meta-analysis on instructor presence (Beege et al., 2023), the eye tracking experiment indicated a strong effect of the visible talking head on various measures of visual attention. In line with previous instructor presence (e.g., van Wermeskerken et al., 2018) and talking head research (Zhang & Yang, 2022), the talking head resulted in learners fixating the overall learning content for a shorter time. In addition, consistent with the findings of van Wermeskerken et al. (2018), learners fixated the newly appearing content for a shorter time and less frequently.

In addition, the present findings were able to expand previous eye tracking research on instructor presence in some respects. Extending the findings of Zhang and Yang (2022), the distracting effect of the talking head was demonstrated also for graphic-based learning content. In addition, Experiment 3 extended the findings of van Wermeskerken et al. (2018) regarding a higher-embodied instructor to the effect that a talking head not only resulted in newly appearing content being fixated for a shorter time and less often, but also more slowly. Hence, the talking head also delayed learners' fixation of newly appearing content, unlike the higher-

embodied instructor in the study by van Wermeskerken et al. (2018). It is possible that the higher-embodied instructor in the study by van Wermeskerken et al. (2018) used gestures to refer to newly appearing content, preventing a delay at this point. Another explanation for this difference might be that the appearance of the content in the video by van Wermeskerken et al. (2018) followed a clear structure (from top to bottom), whereas the video content in my eye tracking study appeared in different places on the slides without a fixed structure. This probably made it more difficult for learners to anticipate where the next content would appear, which could have caused a delay with a visible talking head.

Research Question 5 concerned the extent to which learners' eye movements may explain talking head effects on learning outcomes. Hence, I investigated whether fixation duration of the content could explain the talking head effects on learning, without any evidence of significant mediations. Consistent with studies on high-embodied instructors (van Wermeskerken et al., 2018; Wang, Antonenko, & Dawson, 2020), there were largely no associations between eye movements and learning measures (neither on knowledge test nor on picture recognition task performance). This missing relationship is particularly surprising with regard to picture recognition task performance, because this task explicitly required learners to look at the learning content on the slides, including visual details. As already discussed, the learners seemed to be able to compensate well for the visual distraction caused by the talking head and even limited fixation durations on the content seemed to be sufficient to correctly recognise visual details (cf. Eitel et al., 2012). However, these findings also illustrated the importance of combining process measures such as eye tracking data with other measures, as already discussed in the General Introduction section. If the eye tracking findings were considered in isolation, one would assume a negative effect due to the talking head. However, combining the eye tracking results with different learning measures revealed that this distraction caused by the talking head on the level of process data did not translate into different learning measures, thus qualifying the eye tracking findings.

Nevertheless, it is important to note that across conditions, there was a positive correlation between the number of fixations of the newly appearing content and picture recognition task performance (cf. Appendix C). This relationship at least suggested that visual attention might be associated with learning measures. For future research in this context, the number of fixations might be an interesting measure in addition to fixation duration.

### *3.1.3. Cognitive Load*

In addressing Research Question 4, the experiments also investigated the effects of talking heads on rating measures. In terms of different measures of cognitive load, the findings were also heterogeneous. Experiments 1a, 1b, and 2 recorded the three sources of cognitive load using a scale by Klepsch et al. (2017), while Experiment 3 included one-item ratings of perceived difficulty (as indicator for intrinsic and extraneous cognitive load, although a clear separation was no longer possible here) and invested mental effort (as indicator for germane cognitive load).

With regard to extraneous cognitive load (Research Question 4), the internal meta-analysis across all experiments revealed that the talking head reduced ratings of extraneous cognitive load. When considering the findings at the level of the individual experiments, there was no effect of the talking head in Experiment 1a. However, in Experiments 1b and 2, the talking head consistently led to lower extraneous cognitive load ratings, despite numerous differences between the two experiments (e.g., design, materials). In Experiment 2, reduced extraneous cognitive load ratings induced by the talking head corresponded with reduced ratings of perceived difficulty. In Experiment 3, as in Experiment 1a, no talking head effect on perceived difficulty was observed, consistent with the findings of Ng and Przybylek (2021), where the talking head also did not influence the ratings of extraneous cognitive load. Whereas the pattern is inconsistent and warrants further research, the load-reducing effects in Experiments 1b, 2, and the overall meta-analysis were actually consistent with some other findings from prior research on talking heads (Kizilcec et al., 2015) and instructor presence in general (Wang, Antonenko, Keil, & Dawson, 2020). Similarly, the broad range of findings summarised in the meta-analysis by Beege et al. (2023) supported the reducing effect of a visible instructor on extraneous cognitive load, although it should be noted that this meta-analysis already included the results of Experiments 1a, 1b, and 2 reported in this thesis. The fact that learners expressed lower extraneous cognitive load for videos including a visible talking head might stem from learners spending a lot of time looking at the talking head and therefore being less able to realistically assess the design of the remaining learning content. Hence, learners might tend to primarily evaluate the difficulty of the talking head and not the difficulty of the learning materials presented next to the instructor, presumably not perceiving the talking head as particularly difficult. Beege et al. (2023) provided a similar explanation in their meta-analysis, also attributing the load-reducing effects to the split-attention effect (Ayres & Sweller, 2022) and learners' reduced attention on the content. Accordingly, they point out that social information (i.e., the instructor's visuals) as a fundamental biological process does

not pose a strong cognitive load (Kirschner et al., 2018), explaining the reduced ratings of extraneous cognitive load. However, the missing relationships between fixation duration of the learning content and indicators of cognitive load (i.e., perceived difficulty) in Experiment 3 contradict this explanation. Here, increasing attention to the learning content was not associated with higher ratings of difficulty (cf. Appendix C). However, Experiment 3, unlike the other experiments presented, did not include a detailed assessment of extraneous cognitive load, but only one item on perceived difficulty.

In summary, load-reducing effects of the talking head are not consistent with theory. According to coherence (e.g., Mayer, 2021) or split-attention principle (Ayres & Sweller, 2022), a talking head should, if at all, lead to an increase in extraneous cognitive load. It is conceivable that the learners were not able to assess their actual load accurately using the self-assessment of extraneous cognitive load (Klepsch et al., 2017). It is possible that they did not perceive the design as unfavourable or exhausting, but rather – in line with their overall positive satisfaction ratings in the internal meta-analysis – very favourable, without this necessarily reflecting their actual cognitive load. However, results of explorative mediation analyses from Experiment 2 (cf. Appendix B) did not support this argument by suggesting that learners' ratings of extraneous cognitive load corresponded to their learning outcomes. In particular, the less strenuous a learner assessed videos including a talking head compared to videos not including a talking head, the less detrimental to learning the talking head was compared to videos not including a talking head. Against this background, it would be interesting to investigate talking head effects with other instruments for recording cognitive load, both with other self-reports (e.g., Krieglstein et al., 2023) and with process measures such as pupil dilation (e.g., Lee et al., 2020; Rosch & Vogel-Walcutt, 2013).

With regard to the intrinsic cognitive load, there consistently were no talking head effects in Experiments 1a, 1b, and 2 and, correspondingly, no effects on perceived difficulty ratings in Experiment 3. This pattern of findings is consistent with CLT (e.g., Sweller et al., 1998), which would not suggest any talking head effects on intrinsic cognitive load since adding or removing a talking head should not affect the interactivity of the elements in the learning material.

With respect to germane cognitive load, there were no talking head effects in Experiments 1a and 1b. In Experiment 2, in line with the social agency theory (e.g., Mayer & DaPra, 2012), the visible talking head resulted in higher ratings of germane cognitive load (and also of social presence). In laboratory Experiment 3, there was no systematic recording of

germane cognitive load with three items due to time constraints, but there were no differences in ratings of invested mental effort which was recorded as an indicator for germane cognitive load. Further contributing to answer Research Question 4, the internal meta-analysis across the four experiments supported the mostly absent effects and indicated no overall talking head effect on germane load. Potential reasons for the differing findings between Experiments 2 and 3 could lie in the different measurement instruments, but also in the different settings (online vs. laboratory). To the best of my knowledge, these are the first findings available on the effects of talking head on germane cognitive load, so no real comparison with previous research is possible. The differences between the various studies (between vs. within-subjects factor, online vs. laboratory experiment, different measurement instruments) suggest various potential reasons for different findings. Taken together, most of the experiments and the internal meta-analysis did not support the predictions of social agency theory (Mayer & DaPra, 2012; Mayer et al., 2004), namely an increase in germane cognitive load by adding a social cue, with only Experiment 2 yielding a small to medium effect ($\eta_p^2 = .046$, cf. Cohen, 1988) consistent with theory. Similar to extraneous cognitive load, more research, also using other instruments (e.g., Krieglstein et al., 2023), is desirable.

### *3.1.4. Social Presence*

In addition, Research Question 4 concerned talking head effects on social presence. In this regard, the reported findings were also mixed. Whereas Experiments 1a and 1b showed no effects of talking heads on social presence ratings, both Experiments 2 ($\eta_p^2 = .518$) and Experiments 3 ($\eta_p^2 = .454$) revealed strong positive effects according to Cohen (1988). However, the internal meta-analysis across all four experiments revealed no general talking head effect on social presence ratings. When discussing the deviating effects between Experiments 1a and 1b on the one hand and Experiments 2 and 3 on the other, it should be noted that I used the same three-item scale in all four experiments. It is conceivable that very strong social presence effects only emerged when learners had a direct comparison between formats with and without a talking head as it is the case in the within-subjects designs in Experiments 2 and 3. Since Experiments 2 and 3 involved the same materials, I conducted additional exploratory analyses with these datasets which supported the aforementioned explanation. When considering the social presence ratings for the first video in Experiments 2 and 3, in which the learners had no direct comparison between different formats, there was no significant difference in the ratings between videos with and without a talking head, neither in Experiment 2, $t(110) = 0.66$, $p = .509$, $d = 0.13$, nor in Experiment 3, $t(94) = 0.07$, $p = .945$, $d = 0.01$. However, when considering the social presence ratings of the last video (i.e., the

learners had watched several videos both with and without a talking head), both Experiment 2, $t(99.595) = 5.94$, $p < .001$, $d = 1.12$, and Experiment 3, $t(86.179) = 4.85$, $p < .001$, $d = 0.99$, showed a strong positive effect of the talking head on social presence ratings.[3] However, it is important to note that the findings presented can also be explained by a temporal factor, with social presence ratings, for example, increasing over the course of the experiments. In this respect, different talking head effects over time (especially with longer videos) represent an interesting perspective for future research.

Nonetheless, the aforementioned reasoning is also supported by prior research, which reported no differences in social presence ratings in talking head studies with between-subjects design (Ng & Przybylek, 2021; Rosenthal & Walker, 2020, Study 1; Yuan et al., 2021). In contrast, in line with my findings involving a within-subjects design (Experiments 2 and 3), in Study 2 by Rosenthal & Walker (2020), a talking head enhanced social presence ratings in a repeated-measures design. In the meta-analysis, in which the findings from Experiments 1a, 1b, and 2 were already included, Beege et al. (2023) also identified a positive effect of instructor presence in general on social presence ratings.

Taken together, a talking head may increase the learners' perceived social presence, especially when they have a direct comparison between formats with and without a visible talking head. This is particularly relevant since increased social presence according to the social agency theory constitutes a process leading to deeper processing and improved learning (e.g., Mayer & DaPra, 2012). Despite increased social presence ratings in Experiments 2 and 3 and higher germane cognitive load ratings in Experiment 2, the presence of the talking head did not translate into higher learning outcomes in any of the experiments. In terms of learning measures, however, it should be noted that the social agency theory (e.g., Mayer & DaPra, 2012) primarily assumes higher transfer performance, whereas I only recorded recall of factual information.

### *3.1.5. Other Subjective Ratings*

At this point, I focus on talking head effects on ratings of satisfaction and perceived learning, which were collected in all experiments and were addressed in Research Question 4. For an insight into the findings on other variables (e.g., professionalism, joy of learning), please refer to the individual manuscripts. Findings on satisfaction and perceived learning were heterogeneous in different experiments, with the pattern for both measures varying equally across experiments. Comparable to the findings on germane cognitive load, the talking head

[3] Since variance homogeneity was violated in both comparisons relating to the last videos, the corrected results of the Welch's t-test are reported, yielding deviating degrees of freedom.

did not affect both ratings in Experiments 1a and 1b, whereas the talking head had large (satisfaction, $\eta_p^2 = .252$) or medium (perceived learning, $\eta_p^2 = .079$) positive effects in Experiment 2. Accordingly, learners preferred videos with talking heads in the selection task. However, Experiment 3 showed no differences in satisfaction and perceived learning between videos with and without talking head. Taken together, despite the predominant absence of effects at the experiment level, the internal meta-analysis across the experiments showed a positive effect of the talking head on satisfaction ratings. However, for perceived learning, the meta-analysis revealed no overall talking head effect.

When discussing potential explanations for heterogenous findings in the experiments, the variation of the talking head as a within-subjects factor in Experiment 2 aimed to examine whether differences in rating measures emerged when learners had a direct comparison between formats with and without a talking head. Although the findings from Experiment 2 support this assumption, the absent effects in Experiment 3 suggest that a direct comparison between formats is not a sufficient condition for learners to prefer formats with a talking head over those without. It is conceivable that the within-variation of the talking head in more informal settings (online experiment) had positive effects on evaluation measures, whereas it had no effect in a more formal setting (laboratory experiment), in which the learners' focus probably is less on liking the materials.

The heterogeneous pattern of findings across the experiments reflects the heterogeneous findings in previous research. For instance, the missing talking head effects on ratings of perceived learning and satisfaction in Experiments 1a, 1b, and 3 match the lack of differences in two studies by Rosenthal and Walker (2020) comparing a picture-in-picture and a voiceover format regarding quality and preference ratings in both a between- and a within-subjects design. At the same time, according to the social-cue hypothesis (Colliot & Jamet, 2018), individual findings on talking heads (Kizilcec et al., 2014; Yuan et al., 2021) also indicated positive effects on rating measures in line with Experiment 2 and the internal meta-analysis on satisfaction. In contrast, the body of evidence showing that visible instructors in general (not only talking heads) improved participants' ratings of the learning materials (satisfaction, Wang & Antonenko, 2017; Wang, Antonenko, & Dawson, 2020; perceived learning, Wang & Antonenko, 2017; Wilson et al., 2018) is already much broader and is also supported by comprehensive reviews and meta-analyses (e.g., Alemdag, 2022; Beege et al., 2023; Polat, 2023).

In summary, when answering Research Question 4 regarding talking head effects on satisfaction and perceived learning, the results of the internal meta-analyses indicated that talking heads can positively affect the evaluation of the learning materials in terms of satisfaction, but not in terms of perceived learning. In addition, considering all four experiments, no negative effects of talking heads on learners' ratings could be identified.

#### *3.1.6. Beyond the Talking Head – Other Findings*

The presented experiments also provided interesting findings beyond the talking head effects. Although the additional design features varied in Experiments 1a, 1b, and 2 did not interact with the visibility of the talking head in terms of learning outcomes, Experiment 2 revealed interesting interactions regarding the interplay of slide type and presentation type. In particular, in Experiment 2, which included content-equivalent graphic- and text-based slides, a consistent interaction between slide type and presentation type across various variables (e.g., perceived learning, satisfaction, extraneous cognitive load) emerged. This interaction was reflected in sequential presentation proving beneficial for graphic-based slides as it resulted in higher ratings of satisfaction and perceived learning and lower ratings of extraneous cognitive load. For text-based content, however, the sequential presentation of content had no added value compared to displaying all content at once. This interplay appears reasonable in light of the design of the materials. While the appearance of the content on text slides followed a fixed pattern (bullet points appearing stepwise from top to bottom), the content on the graphic slides in the videos appeared without a uniform structure. Hence, the stepwise fade-in may have supported learners' orientation on the slides and thus reduced their extraneous processing (cf. signaling principle, van Gog, 2022). In contrast, Experiment 1b also varied whether graphic-based and text-based content appeared with sequential or static presentation and revealed no corresponding findings. However, this experiment involved different learning materials with the graphic-based and text-based videos covering different topics.

### 3.2. Theoretical Implications

Some of the points mentioned above already indicated that the overall findings also have implications for the underlying theories. The findings presented largely support the image principle (Mayer, 2021). With regard to learning measures, the inclusion of the talking head had no added value at any point, and in some cases even had a small negative effect. Also following the argumentation of the image principle, the talking head as a low-embodied instructor distracted from the learning content on the level of process data measuring learners' eye movements (Experiment 3).

With regard to support for the coherence principle (e.g., Mayer, 2021) applying to talking heads, only the small detrimental effect on learning in Experiment 2 can be mentioned. Beyond this, the findings provide little evidence in favour of excluding the talking head as an interesting but irrelevant element of the learning material. Although the talking head can be characterised as an interesting element in light of the strong visual distraction in Experiment 3, the visible instructor did not lead to an increase in extraneous cognitive load in any of the experiments. On the contrary, the internal meta-analysis demonstrated a decrease in learners' self-reported extraneous cognitive load due to the talking head, which might reflect the fact that learners evaluated the talking head rather than the remaining learning materials. However, the learning materials next to the talking head showed a comparatively low level of element interactivity because they involved mainly facts, and the individual short videos were not linked to each other. It could be interesting for future research to investigate talking head effects next to learning materials with a higher complexity (and also higher element interactivity). It is possible that a talking head leads to learning losses next to such materials.

In addition, the present experiments did not provide a clear answer with regard to the question of whether a talking head can be regarded as a seductive detail (Garner et al., 1989). While the eye tracking findings clearly suggest that the talking head is an interesting element for the learners, it still needs to be clarified whether a talking head is really an irrelevant feature, especially against the background of the high social presence ratings in Experiments 2 and 3. It is conceivable that, regardless of the visual learning materials, the mere fact that learners can see someone presenting the contents brings added value while listening because the talking head may help learners to keep attention on the spoken narration. This would require systematically testing content that was only covered by the spoken narration to determine whether learners remember this content better with a talking head. Further, it is also conceivable that certain elements of the talking head's face (e.g., facial expressions, lip movements) may provide relevant information. This reasoning is supported by the findings of Võ et al. (2012), who observed that when looking at a face, individuals' fixations shifted to the mouth when they heard an audio speech track. However, it should be considered that elements such as lip movements may be more or less relevant depending on a talking head's size in the learning material, with very small talking heads not allowing learners to recognise these movements properly.

In view of cognitive load findings, the experiments also provided no evidence for the split-attention principle (Ayres & Sweller, 2022), since the talking head had an overall reducing effect on extraneous cognitive load. Nevertheless, as for the coherence principle (e.g., Mayer,

2021), the small learning-impeding effect from Experiment 2 may serve as indication for a split-attention effect. Similarly, the eye tracking findings from Experiment 3, namely that learners fixated overall learning content for a shorter time and newly appearing content for a shorter time, less frequently, and more slowly, support the idea that learners split their attention due to the presence of a talking head. However, it may be discussed whether the classic split-attention principle applies to talking heads at all. It is unclear whether the talking head is considered a relevant source of information, since learners did not necessarily need to integrate the visual talking head information with the remaining learning materials. However, my eye tracking findings provided evidence that participants empirically split their attention, which may be even more harmful than classic split-attention in view of the rather limited information value of the talking head. Nonetheless, the findings of Experiment 3 offered only very limited evidence that the reduced attention to the learning content had any detrimental effect on learning. Overall, mediation analyses showed no indications for the fixation duration of the learning content contributing to explain talking head effects on learning measures (i.e., knowledge test and picture recognition task performance). Only a small (according to Cohen, 1988) positive correlation ($r$ = .236) between the number of fixations of the newly appearing content and picture recognition task performance across conditions suggested that increased attention to the newly appearing learning content may be positively related to the recognition of visual content from the videos.

The reported findings only partially supported the social agency theory (e.g., Mayer & DaPra, 2012). In two experiments (Experiments 2 and 3), strong positive effects of the talking head as social cue on social presence ratings were evident, in line with the social agency theory. In one investigation (Experiment 2), there were also indications of deeper processing (higher germane cognitive load) due to the talking head. However, there were no learning-promoting talking head effects at any point, not even in the experiment revealing higher social presence and higher germane cognitive load ratings due to the talking head (Experiment 2). On the contrary, I observed a rather small learning-hindering effect in the respective experiment. This picture appeared not only when considering the main effects of Experiment 2, but also in an exploratory mediation analysis, which showed that social presence ratings did not explain the talking head effect on learning outcomes. However, it is important to note that, according to the social agency theory (e.g., Mayer & DaPra, 2012), learning-enhancing effects should primarily concern transfer performance, which was not investigated in the reported experiments. This is because their focus was on determining whether the talking head impacts the extraction of factual information from the learning material. Nonetheless, the presented experiments

indicated that the processes postulated by the social agency theory (i.e., social cues → higher social presence → deeper processing → improved learning) cannot be considered to be automated processes.

Since the studies did not show any learning-promoting effects of the visible talking head, which is characterised by a rather low embodiment, the findings are generally in line with the embodiment principle (e.g., Fiorella, 2022). At the same time, however, it is important to emphasise that the instructors' embodiment was not systematically varied in any of the experiments, implying that systematic conclusions are not possible in this regard.

The social-cue hypothesis (Colliot & Jamet, 2018), which assumes that social cues may also contribute to a better learning experience, was supported by the results of the internal meta-analysis on satisfaction, which showed that the talking head enhanced learners' satisfaction. Further, particularly Experiment 2 provided support for this idea due to the positive talking head effects on various ratings (e.g., perceived learning). Aside from this, however, there were largely absent effects on various ratings in the remaining experiments, which is also supported by the internal meta-analysis showing no general talking head effect on perceived learning. This suggests that there may be boundary conditions or moderating factors (e.g., study setting) for talking head effects on learners' affective reactions.

In view of the internal meta-analysis across the experiments showing no talking head effects on learning outcomes, it is also conceivable that the effects postulated by the different theories cancelled each other out. Hence, the talking head's potentially learning-inhibiting effect, proposed by the coherence (e.g., Mayer, 2021) and split-attention principles (Ayres & Sweller, 2022), may have been neutralised by the social processes postulated by the social agency theory (e.g., Mayer & DaPra, 2012). While a talking head is not considered to be of particular benefit according to the embodiment principle (e.g., Fiorella, 2022), the social agency theory refers to social cues in general and the high social presence ratings in Experiments 2 and 3 at least indicated that social processes may be involved when using talking heads despite their low embodiment. However, it should be noted that the data at the meta-analyses level did not support the underlying processes postulated in each case. In particular, I observed no increased extraneous load through the violation of coherence or split-attention principle and no increased germane cognitive load due to a social cue.

Beyond specific theories, the question arises as to what extent the heterogeneous findings on talking heads in my experiments and in research in general may be moderated not only by characteristics of the learning material (different types of instructors, different types of

learning materials), but also by learners' characteristics. I already made the first attempt in this regard by trying to include learners' individual gaze behaviour as an explanatory factor in Experiment 3. Although my findings did not provide evidence in this regard, research already revealed initial indications that the talking head might have different effects for different individuals. For instance, in the study by Kokoç et al. (2020), the learning-enhancing effect of a talking head was higher when learners' sustained attention was high compared to when it was low. In light of this, it would be useful to increasingly include different individual prerequisites beyond prior knowledge (e.g., mind-wandering tendency, sustained attention, working memory capacity) in theories. This topic is also particularly relevant because technical progress offers more and more opportunities to adapt learning materials to different learners.

### 3.3. Strengths

This thesis and the individual experiments have numerous strengths, but also limitations, which I will examine in more detail in the following sections. First of all, it should be emphasised that the effects of talking heads were investigated in various experiments in different contexts (online experiments, laboratory experiments) using different designs (between- and within-variation). At this point, it is particularly important to emphasise that this thesis can draw on findings from both an online setting and a laboratory context. Due to the COVID-19 pandemic, online studies were the method of choice, with the online setting being characterised by high ecological validity in relation to the research objective (i.e., educational online videos). Nonetheless, the addition of laboratory data in Experiment 3 constitutes a particular strength, especially as the use of process data systematically expanded findings of Experiments 1a, 1b, and 2, and allowed conclusions about learners' attentional processes. In addition, data from a laboratory context is characterised by a higher internal validity and the laboratory context allowed more control over whether the learners were watching the videos as intended. Although I also tried to achieve maximum control in the online context by using targeted control questions, the laboratory environment offers a higher level of control and therefore a better data quality.

Further, the experiments addressed different potential influencing factors (slide type, presentation type, learners' eye movements) and also covered a wide range of different dependent variables. These included learning outcomes with knowledge acquisition (measured via multiple choice and free recall) and picture recognition, process data (fixation duration, number of fixations, time to first fixation), measures of cognitive load, various subjective rating measures, and a selection task. This systematically expanded the existing talking head research

at various points, for instance by addressing the fixation duration of the content as potential explanatory mechanism for learning outcomes.

A further strength of this thesis certainly lies in the strong experimental design in the individual studies. Most importantly, the applied designs provided the opportunity to systematically investigate the effects of a talking head and other design features (e.g., sequential presentation) across different topics. In many studies on instructor presence in general (e.g., van Wermeskerken et al., 2018) and talking heads in particular (e.g., Kizilcec et al., 2014), the studies only examined single videos on single topics, so that the findings can be generalised to different topics only to a limited extent. Hence, the different video topics from different subject areas in my experiments were intended to increase the generalisability of the findings with regard to this aspect. At the same time, systematic manipulation and balancing prevented a confound between topics and condition. Nonetheless, the strong experimental design also entailed potential disadvantages, which I will address separately in the next sections.

### 3.4. Limitations and Directions for Future Research

When interpreting the findings, some limitations must also be taken into account, which also give rise to questions for future research. With regard to the learning measures recorded, it should be noted that none of the experiments recorded higher order learning outcomes such as transfer. It would have been particularly interesting to record transfer learning against the background of the social agency theory (e.g., Mayer & DaPra, 2012), since the social agency theory postulates deeper processing and higher transfer learning triggered by social cues and social presence. However, the approach of this thesis was to identify potential boundary conditions and disadvantages of talking heads, which were hardly shown in previous research despite the strong influence on visual attention (i.e., distraction from visual learning content). A distracting effect of the talking head on a visual level (e.g., by missing certain content or details) should already be evident when extracting facts from the learning materials (i.e., information selection). Although this may have implications for higher order learning, since missed information cannot be further processed, recording higher order learning measures was beyond the scope of this research. One challenge for future research consists in finding a way to ensure that the talking head does not interfere with the extraction of information, but promotes transfer learning.

Another reason for the restriction to the recording of retention and to the use of materials with rather low element interactivity lies in the comparatively highly controlled study designs. The use of videos on four to eight different topics and the balancing of topics and (talking head)

conditions led to the fact that the individual independent videos were relatively short (1 to 5 minutes). Consequently, it was difficult to capture transfer learning as the short video duration limited the ability to provide in-depth conceptual contents regarding the individual topics.

In general, the video length might be considered as a potential limitation resulting from methodological choices. The individual videos in Experiments 1a and 1b lasted between 2.5 and 5 minutes, and in Experiments 2 and 3 each video lasted around 1 minute. This implies that the findings cannot be automatically generalised to longer videos (e.g., lecture videos that last 1.5 hours). However, it should be noted that the use of shorter videos is not necessarily ecologically invalid. On the contrary, some research suggested that an educational video should ideally last no longer than 6 minutes (Guo et al., 2014). Further, the decision to use short videos is based on the fact that this allows the examination of videos on different topics with a reasonable study duration (including eye tracking). In addition, the use of several short videos enables more measurements, which is reflected in higher power and higher reliability. Nonetheless, it is the task of future research to test whether the video duration could serve as a potential moderator for talking head effects. Hence, it might be interesting to look at talking head effects in different sequences of longer videos. On the one hand, it would be conceivable that talking head effects decrease with increasing video duration (i.e., habituation effect). On the other hand, in light of the findings of Guo et al. (2014) described above, it is also conceivable that beneficial talking head effects may only come into play with increasing video duration. It is conceivable that a talking head only develops added value, for instance in terms of motivation, once a video has reached a certain length (e.g., 6 minutes, Guo et al., 2014).

In terms of the recorded learning outcomes, it should also be noted that the experiments only captured short-term retention directly after the learning phase and not medium- or long-term retention. Investigating talking head effects on medium- and long-term retention represents a potential perspective for future research. In this context, it would also be interesting to explore whether a talking head might even have added value if it is visible for learners when they retrieve information, in the sense of a retrieval cue. For instance, a previous study on seductive details (Schneider et al., 2020) showed that participants performed better in a knowledge test when decorative pictures from the learning material were also visible as a retrieval cue in the testing phase.

In addition, many variables in the experiments, including cognitive load, were recorded using self-report measures. Although this is certainly a reasonable approach for numerous rating aspects in order to determine how learners like different videos, further measures would be

desirable, especially for recording cognitive load. For instance, it would be interesting for future research to use new scales (e.g., Krieglstein et al., 2023) or pupil dilation (e.g., Lee et al., 2020; Rosch & Vogel-Walcutt, 2013) as measures of cognitive load. However, when utilising pupil dilation, it is important to consider that the talking head might change the brightness of the image, and this should be taken into account when designing the materials. Further, it is important to note that pupil dilation may not allow conclusions on different types of cognitive load. Hence, as with other eye tracking metrics, combining pupil dilation with additional measures (e.g. self-reports) can be useful.

Similarly, the learners also self-assessed their prior knowledge in order to rule out systematic pre-existing differences between conditions. However, this self-report may not necessarily always correspond to their actual (prior) knowledge (e.g., Kardas and O'Brien, 2018; Toftness et al., 2018). However, I decided against assessing learners' prior knowledge with a knowledge test to avoid drawing the learners' attention to important learning content in advance (cf. Pan & Carpenter, 2023). In addition, I tried to minimise the role of prior knowledge as much as possible in all experiments by using video topics for which most people presumably had little or no prior knowledge (e.g., Burkina Faso), which is also supported by the low prior knowledge scores in all experiments. In addition, it should be noted that, in contrast to cognitive load and social presence, certain variables (satisfaction, perceived learning) were only recorded with one item each, entailing the problem that these one-item measures do not allow the calculation of reliability measures.

Further, a talking head's characteristics (e.g., age, gender, attractiveness, voice, similarity to learners, personal familiarity) could impact different results. In terms of familiarity, I controlled for this in all experiments by excluding all participants who stated that they knew the talking head. I also kept the talking head constant within experiments to avoid systematic confounds. However, it should be noted that the talking head was not constant across all experiments, but only in Experiments 1a and 1b as well as in Experiments 2 and 3. Therefore, it cannot be ruled out that different talking heads (same age, different gender) also contributed to explaining potential differences between Experiments 1a and 1b on the one hand and Experiments 2 and 3 on the other hand. However, research by Hoogerheide et al. (2018) and Schrader et al. (2021) revealed that the instructor's gender had no effect on learning outcomes. However, it should be noted at this point that it was beyond the scope of this work to examine various talking head characteristics systemically. One reason for this is that instructors who produce educational videos themselves or have to decide whether to show themselves in

lectures are not able to change the majority of these parameters and I aimed to focus on aspects that can be controlled by the instructors.

With regard to learning materials that appeared next to the talking head, I consistently used videos with narrated slides (i.e., PowerPoint slides). This in turn implies that the results cannot simply be generalised to all types of videos. However, it is important to note that educational videos and online lectures including a talking head mostly use slides or similar presentations (e.g., Polat, 2023). Further, many other educational videos also use content appearing step-by-step, for instance Studyflix (https://studyflix.de/). Nonetheless, with regard to the materials next to the talking head, it would also be interesting to examine more complex materials that build on each other rather than just presenting facts one after the other. It is conceivable that the distraction caused by the talking head may not have been very problematic in the materials examined here (including graphic-based materials). Distraction due to a talking head might be more detrimental with content containing information that builds on each other than with content presenting isolated facts.

In addition, the use of both online and laboratory experiments made it possible to investigate whether talking head effects are equally evident in different contextual factors. However, I did not systematically vary these context factors within a single experiment, meaning that no causal conclusions can be drawn at this point and different findings between the experiments only suggest that the setting could be partly responsible, because the experiments varied regarding different features (e.g., sample, use of eye tracking). Nonetheless, contextual factors of online and laboratory experiments are difficult to manipulate, or only in a very isolated way.

Although the online experiments described are characterised by a comparatively high ecological validity, watching educational videos in the context of an online study does not equate to online learning, for instance in MOOCs or on YouTube. In such cases, the analysis of log data could serve to provide even better insights into learners' natural usage behaviour compared to experiments, which was already successfully realised in other contexts (e.g., Anders et al., 2024; Merkt et al., 2022). However, such learning analytics also pose certain challenges due to the lack of control, for example with regard to the possible missing content equivalence resulting from different usage patterns. Nonetheless, usage data might be employed to determine, for example, whether learners tend to continue watching videos with a talking head rather than videos without a talking head (cf. Guo et al., 2014).

### 3.5. Practical Implications

Overall, talking heads had no general effect on learning, despite of a small detrimental effect on learning in a less formalised environment, independent of other video characteristics in Experiment 2. In addition, the talking head enhanced learners' satisfaction and decreased their self-reported extraneous cognitive load, but did not affect other ratings (i.e., social presence, perceived learning, germane cognitive load). At no point did the talking head have a negative effect on the subjective ratings of the materials. Hence, in light of the overall meta-analyses, there is no reason not to include a talking head in an educational video. Although the inclusion of a talking head did not increase learning outcomes based on the present findings, it can improve the learners' learning experience and reduce their self-reported extraneous cognitive load. If, for instance, lecturers wish to improve their learners' satisfaction, it may be advisable to switch their camera on rather than off.

However, the findings from Experiment 2 suggested that learning measures may have different outcomes than learners' ratings, in that the talking head led to poorer learning outcomes, but to higher satisfaction and perceived learning ratings and learners' preference in a selection task. Consequently, design features that contribute to popularity are not necessarily beneficial for learning, which can prove problematic, especially in informal learning contexts. In this context, not liking a video may lead to a systematic dropout, which manifests itself in learners not watching or continuing to watch unpopular videos in real free choice learning scenarios, even though they may be designed to be beneficial for learning. Accordingly, in a study by Wilson et al. (2018, Experiment 3), learners indicated a lower likelihood of dropping a course and a higher likelihood of completing a lecture with a present instructor. That is of particular interest because in the same set of experiments, Wilson et al. (2018) observed that lectures with a visible instructor resulted in worse learning outcomes than the comparable lecture without an instructor. Hence, when deciding to include or exclude a talking head, practitioners may have to consider a trade-off between popularity and educational use in some cases. Further research is required to understand why and under what circumstances talking head effects can sometimes hinder learning, so that practitioners do not have to choose between educational use and popularity in the long term.

In this respect, promising perspectives arise for research and practice in the distant future. Once broader evidence is available for favourable and unfavourable conditions of talking heads, the use of adaptive talking heads would be an interesting long-term prospect for research and practice. For instance, it would be conceivable to adaptively fade out a talking

head in educational videos as soon as learners look at the talking head for too long. In line with this, previous research by D'Mello et al. (2012) investigated a gaze-reactive tutoring system which monitored learners' eye movements in order to prevent them from drifting off. The resulting prompts from the instructor led learners to refocus on the learning content and improved their learning. Since previous research already demonstrated that increasing attention on a talking head is related to mind-wandering (Zhang et al., 2020), an adaptive fade-in and fade-out of a talking head depending on the learners' eye movements in the sense of a gaze-reactive educational video might reflect a beneficial perspective.

### 3.6. Conclusion

In summary, even though the talking head had a learning-hindering effect in one of four experiments, the internal meta-analysis provided no evidence for a general interference of the talking head with learning across all experiments. Although the talking head exhibited a potentially distracting effect at the level of process data measuring learners' eye movements, this distraction did not translate into impaired learning outcomes. Nevertheless, when considering the level of learning outcomes, future research needs to continue identifying the exact conditions under which talking heads may have detrimental effects on learning in order to be able to specifically avoid these conditions in practice.

However, when considering how learners evaluated the presence of a talking head, its inclusion contributed to the overall satisfaction of the learners and reduced their self-reported extraneous cognitive load. Additionally, the talking head did not negatively affect the learners' subjective affective ratings in any of the experiments. Thus, on the level of learners' ratings, lecturers should be recommended to switch on the camera in order to improve their students' learning experience. Therefore, taking into account these different aspects, designers of educational videos need to consider their priorities when deciding in favour or against using a talking head in their educational videos.

## References

Abrams, R. A., & Christ, S. E. (2003). Motion onset captures attention. *Psychological Science*, *14*(5), 427–432. https://doi.org/10.1111/1467-9280.01458

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

Anders, G., Buder, J., Merkt, M., Egger, E., & Huff, M. (2024). Associations between mind wandering, viewer interactions, and the meaningful structure of educational videos. *Computers & Education*, *212*, Article 104996. https://doi.org/10.1016/j.compedu.2024.104996

Ayres, P., & Sweller, J. (2022). The Split-Attention Principle in Multimedia Learning. In R. E. Mayer & L. Fiorella (Eds.), *The Cambridge handbook of multimedia learning* (3rd ed., pp. 199–211). Cambridge University Press. https://doi.org/10.1017/9781108894333.020

Baber, H. (2022). Social interaction and effectiveness of the online learning – A moderating role of maintaining social distance during the pandemic COVID-19. *Asian Education and Development Studies*, *11*(1), 159–171. https://doi.org/10.1108/AEDS-09-2020-0209

Baddeley, A. (1992). Working memory. *Science*, *255*(5044), 556–559. https://doi.org/10.1126/science.1736359

Beege, M., Ninaus, M., Schneider, S., Nebel, S., Schlemmel, J., Weidenmüller, J., Moeller, K., & Rey, G. D. (2020). Investigating the effects of beat and deictic gestures of a lecturer in educational videos. *Computers & Education*, *156*, Article 103955. https://doi.org/10.1016/j.compedu.2020.103955

Beege, M., Schneider, S., Nebel, S., & Rey, G. D. (2017). Look into my eyes! Exploring the effect of addressing in educational videos. *Learning and Instruction*, *49*, 113–120. https://doi.org/10.1016/j.learninstruc.2017.01.004

Beege, M., Schroeder, N. L., Heidig, S., Rey, G. D., & Schneider, S. (2023). The instructor presence effect and its moderators in instructional video: A series of meta-analyses. *Educational Research Review*, *41*, Article 100564. https://doi.org/10.1016/j.edurev.2023.100564

Borenstein, M., Hedges, L. V., Higgins, J. P. T., & Rothstein, H. R. (2009). *Introduction to Meta-Analysis*. Wiley. https://doi.org/10.1002/9780470743386

Chen, C.-M., & Wu, C.-H. (2015). Effects of different video lecture types on sustained attention, emotion, cognitive load, and learning performance. *Computers & Education*, *80*, 108–121. https://doi.org/10.1016/j.compedu.2014.08.015

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*. Lawrence Erlbaum Associates.

Colliot, T., & Jamet, É. (2018). Understanding the effects of a teacher video on learning from a multimedia document: an eye-tracking study. *Educational Technology Research and Development*, *66*(6), 1415–1433. https://doi.org/10.1007/s11423-018-9594-x

Coursera. (2023, October 26). *Coursera Reports Third Quarter 2023 Financial Results.* https://investor.coursera.com/news/news-details/2023/Coursera-Reports-Third-Quarter-2023-Financial-Results/default.aspx

Dargue, N., Sweller, N., & Jones, M. P. (2019). When our hands help us understand: A meta-analysis into the effects of gesture on comprehension. *Psychological Bulletin*, *145*(8), 765–784. https://doi.org/10.1037/bul0000202

D'Mello, S., Olney, A., Williams, C., & Hays, P. (2012). Gaze tutor: A gaze-reactive intelligent tutoring system. *International Journal of Human-Computer Studies*, *70*(5), 377–398. https://doi.org/10.1016/j.ijhcs.2012.01.004

Eitel, A., Scheiter, K., & Schüler, A. (2012). The time course of information extraction from instructional diagrams. *Perceptual and Motor Skills*, *115*(3), 677–701. https://doi.org/10.2466/22.23.PMS.115.6.677-701

Feierabend, S., & Klinger, W. (2003). *Lehrer/-Innen und Medien 2003*. Medienpädagogischer Forschungsverbund Südwest.

Fiorella, L. (2022). The Embodiment Principle in Multimedia Learning. In R. E. Mayer & L. Fiorella (Eds.), *The Cambridge handbook of multimedia learning* (3rd ed., pp. 286–295). Cambridge University Press. https://doi.org/10.1017/9781108894333.030

Fiorella, L., & Mayer, R. E. (2022a). Principles Based on Social Cues in Multimedia Learning. In R. E. Mayer & L. Fiorella (Eds.), *The Cambridge handbook of multimedia learning* (3rd ed., pp. 277–285). Cambridge University Press. https://doi.org/10.1017/9781108894333.029

Fiorella, L., & Mayer, R. E. (2022b). Principles for Reducing Extraneous Processing in Multimedia Learning. In R. E. Mayer & L. Fiorella (Eds.), *The Cambridge handbook of multimedia learning* (3rd ed., pp. 185–198). Cambridge University Press. https://doi.org/10.1017/9781108894333.019

Fiorella, L., Stull, A. T., Kuhlmann, S., & Mayer, R. E. (2019). Instructor presence in video lectures: The role of dynamic drawings, eye contact, and instructor visibility. *Journal of Educational Psychology*, *111*(7), 1162–1171. https://doi.org/10.1037/edu0000325

Garner, R., Gillingham, M. G., & White, C. S. (1989). Effects of 'Seductive Details' on Macroprocessing and Microprocessing in Adults and Children. *Cognition and Instruction*, *6*(1), 41–57. https://doi.org/10.1207/s1532690xci0601_2

Ginns, P. (2006). Integrating information: A meta-analysis of the spatial contiguity and temporal contiguity effects. *Learning and Instruction*, *16*(6), 511–525. https://doi.org/10.1016/j.learninstruc.2006.10.001

Glaser, M., & Schwan, S. (2015). Explaining pictures: How verbal cues influence processing of pictorial learning material. *Journal of Educational Psychology*, *107*(4), 1006–1018. https://doi.org/10.1037/edu0000044

Goulet-Pelletier, J.-C., & Cousineau, D. (2018). A review of effect sizes and their confidence intervals, Part I: The Cohen's d family. *The Quantitative Methods for Psychology*, *14*(4), 242–265. https://doi.org/10.20982/tqmp.14.4.p242

Gullberg, M., & Holmqvist, K. (2006). What speakers do and what addressees look at. *Pragmatics & Cognition*, *14*(1), 53–82. https://doi.org/10.1075/pc.14.1.05gul

Guo, P. J., Kim, J., & Rubin, R. (2014). How video production affects student engagement. In M. Sahami, A. Fox, M. A. Hearst, & M. T. Chi (Eds.), *Proceedings of the first ACM conference on Learning @ scale conference - L@S '14* (pp. 41–50). Association for Computing Machinery. https://doi.org/10.1145/2556325.2566239

Hansch, A., Hillers, L., McConachie, K., Newman, C., Schildhauer, T., & Schmidt, P. (2015). Video and online learning: Critical reflections and findings from the field. *SSRN Electronic Journal*. Advance online publication. https://doi.org/10.2139/ssrn.2577882

Henderson, J. M., Weeks, P. A., & Hollingworth, A. (1999). The effects of semantic consistency on eye movements during complex scene viewing. *Journal of Experimental Psychology: Human Perception and Performance*, *25*(1), 210–228. https://doi.org/10.1037/0096-1523.25.1.210

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

Hew, K. F., & Lo, C. K. (2018). Flipped classroom improves student learning in health professions education: A meta-analysis. *BMC Medical Education*, *18*(1), 38. https://doi.org/10.1186/s12909-018-1144-z

Hew, K. F., & Lo, C. K. (2020). Comparing video styles and study strategies during video-recorded lectures: effects on secondary school mathematics students' preference and learning. *Interactive Learning Environments*, *28*(7), 847–864. https://doi.org/10.1080/10494820.2018.1545671

Holmqvist, K., & Andersson, R. (2017). *Eye tracking: A comprehensive guide to methods, paradigms, and measures* (2nd edition). Lund Eye-Tracking Research Institute.

Homer, B. D., Plass, J. L., & Blake, L. (2008). The effects of video on cognitive load and social presence in multimedia-learning. *Computers in Human Behavior*, *24*(3), 786–797. https://doi.org/10.1016/j.chb.2007.02.009

Hong, J., Pi, Z., & Yang, J. (2018). Learning declarative and procedural knowledge via video lectures: Cognitive load and learning effectiveness. *Innovations in Education and Teaching International*, *55*(1), 74–81. https://doi.org/10.1080/14703297.2016.1237371

Hoogerheide, V., van Wermeskerken, M., van Nassau, H., & van Gog, T. (2018). Model-observer similarity and task-appropriateness in learning from video modeling examples: Do model and student gender affect test performance, self-efficacy, and perceived competence? *Computers in Human Behavior*, *89*, 457–464. https://doi.org/10.1016/j.chb.2017.11.012

Johnson, M. H., Dziurawiec, S., Ellis, H., & Morton, J. (1991). Newborns' preferential tracking of face-like stimuli and its subsequent decline. *Cognition*, *40*(1-2), 1–19. https://doi.org/10.1016/0010-0277(91)90045-6

Just, M. A., & Carpenter, P. A. (1976). Eye fixations and cognitive processes. *Cognitive Psychology*, *8*(4), 441–480. https://doi.org/10.1016/0010-0285(76)90015-3

Just, M. A., & Carpenter, P. A. (1980). A theory of reading: From eye fixations to comprehension. *Psychological Review*, *87*(4), 329–354. https://doi.org/10.1037/0033-295X.87.4.329

Kardas, M., & O'Brien, E. (2018). Easier seen than done: Merely watching others perform can foster an illusion of skill acquisition. *Psychological Science*, *29*(4), 521–536. https://doi.org/10.1177/0956797617740646

Kay, R. H. (2012). Exploring the use of video podcasts in education: A comprehensive review of the literature. *Computers in Human Behavior*, *28*(3), 820–831. https://doi.org/10.1016/j.chb.2012.01.011

Kelley, G. A., & Kelley, K. S. (2012). Statistical models for meta-analysis: A brief tutorial. *World Journal of Methodology*, *2*(4), 27–32. https://doi.org/10.5662/wjm.v2.i4.27

Kirschner, P. A., Sweller, J., Kirschner, F., & Zambrano R, J. (2018). From Cognitive Load Theory to Collaborative Cognitive Load Theory. *International Journal of Computer-Supported Collaborative Learning*, *13*(2), 213–233. https://doi.org/10.1007/s11412-018-9277-y

Kizilcec, R. F., Bailenson, J. N., & Gomez, C. J. (2015). The instructor's face in video instruction: Evidence from two large-scale field studies. *Journal of Educational Psychology*, *107*(3), 724–739. https://doi.org/10.1037/edu0000013

Kizilcec, R. F., Papadopoulos, K., & Sritanyaratana, L. (2014). Showing face in video instruction: Effects on information retention, visual attention, and affect. In M. Jones, P. Palanque, A. Schmidt, & T. Grossman (Eds.), *Proceedings of the 32nd annual ACM conference on Human factors in computing systems - CHI '14* (pp. 2095–2102). Association for Computing Machinery. https://doi.org/10.1145/2556288.2557207

Klepsch, M., Schmitz, F., & Seufert, T. (2017). Development and validation of two instruments measuring intrinsic, extraneous, and germane cognitive load. *Frontiers in Psychology*, *8*, Article 1997. https://doi.org/10.3389/fpsyg.2017.01997

Kokoç, M., IIgaz, H., & Altun, A. (2020). Effects of sustained attention and video lecture types on learning performances. *Educational Technology Research and Development*, *68*(6), 3015–3039. https://doi.org/10.1007/s11423-020-09829-7

Krasich, K., McManus, R., Hutt, S., Faber, M., D'Mello, S. K., & Brockmole, J. R. (2018). Gaze-based signatures of mind wandering during real-world scene processing. *Journal*

*of Experimental Psychology. General*, *147*(8), 1111–1124. https://doi.org/10.1037/xge0000411

Krieglstein, F., Beege, M., Rey, G. D., Sanchez-Stockhammer, C., & Schneider, S. (2023). Development and Validation of a Theory-Based Questionnaire to Measure Different Types of Cognitive Load. *Educational Psychology Review*, *35*(1). https://doi.org/10.1007/s10648-023-09738-0

Kuang, Z., Wang, F., Xie, H., Mayer, R. E., & Hu, X. (2023). Effect of the Instructor's Eye Gaze on Student Learning from Video Lectures: Evidence from Two Three-Level Meta-Analyses. *Educational Psychology Review*, *35*(4). https://doi.org/10.1007/s10648-023-09820-7

Lee, J. Y., Donkers, J., Jarodzka, H., Sellenraad, G., & van Merriënboer, J. J. (2020). Different effects of pausing on cognitive load in a medical simulation game. *Computers in Human Behavior*, *110*, Article 106385. https://doi.org/10.1016/j.chb.2020.106385

Mayer, R. E. (2021). *Multimedia learning* (3rd ed.). Cambridge University Press.

Mayer, R. E. (2022). Cognitive Theory of Multimedia Learning. In R. E. Mayer & L. Fiorella (Eds.), *The Cambridge handbook of multimedia learning* (3rd ed., pp. 57–72). Cambridge University Press. https://doi.org/10.1017/9781108894333.008

Mayer, R. E., & DaPra, C. S. (2012). An embodiment effect in computer-based learning with animated pedagogical agents. *Journal of Experimental Psychology. Applied*, *18*(3), 239–252. https://doi.org/10.1037/a0028616

Mayer, R. E., & Fiorella, L. (2014). Principles for reducing extraneous processing in multimedia learning: Coherence, signaling, redundancy, spatial contiguity, and temporal contiguity principles. In R. E. Mayer (Ed.), *The Cambridge handbook of multimedia learning* (2nd ed., pp. 279–315). Cambridge University Press.

Mayer, R. E., & Fiorella, L. (2022). Introduction to Multimedia Learning. In R. E. Mayer & L. Fiorella (Eds.), *The Cambridge handbook of multimedia learning* (3rd ed., pp. 3–16). Cambridge University Press. https://doi.org/10.1017/9781108894333.003

Mayer, R. E., & Moreno, R. (2003). Nine ways to reduce cognitive load in multimedia learning. *Educational Psychologist*, *38*(1), 43–52. https://doi.org/10.1207/S15326985EP3801_6

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*(2), 389–395. https://doi.org/10.1037/0022-0663.96.2.389

Meier, J., Jong, B. de, van Montfort, D. P., Verdonschot, A., van Wermeskerken, M., & van Gog, T. (2023). Do social cues in instructional videos affect attention allocation, perceived cognitive load, and learning outcomes under different visual complexity conditions? *Journal of Computer Assisted Learning*, *39*(4), 1339–1353. https://doi.org/10.1111/jcal.12803

Merkt, M., Hoppe, A., Bruns, G., Ewerth, R., & Huff, M. (2022). Pushing the button: Why do learners pause online videos? *Computers & Education*, *176*, Article 104355. https://doi.org/10.1016/j.compedu.2021.104355

Miller, G. A. (1956). The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological Review*, *63*(2), 81–97. https://doi.org/10.1037/h0043158

Moreno, R. (2007). Optimising learning from animations by minimising cognitive load: cognitive and affective consequences of signalling and segmentation methods. *Applied Cognitive Psychology*, *21*(6), 765–781. https://doi.org/10.1002/acp.1348

Moreno, R., Mayer, R. E., Spires, H. A., & Lester, J. C. (2001). The Case for Social Agency in Computer-Based Teaching: Do Students Learn More Deeply When They Interact With Animated Pedagogical Agents? *Cognition and Instruction*, *19*(2), 177–213. https://doi.org/10.1207/S1532690XCI1902_02

Ng, Y. Y., & Przybylek, A. (2021). Instructor Presence in Video Lectures: Preliminary Findings From an Online Experiment. *IEEE Access*, *9*, 36485–36499. https://doi.org/10.1109/ACCESS.2021.3058735

O'Neil-Hart, C. (2017). *Why you should lean into how-to content in 2018*. Think with Google. https://www.thinkwithgoogle.com/_qs/documents/4703/1145-whats-missing-2018-strategy-Download.pdf

Paivio, A. (1990). *Mental Representations: A dual coding approach*. Oxford University Press. https://doi.org/10.1093/acprof:oso/9780195066661.001.0001

Pan, S. C., & Carpenter, S. K. (2023). Prequestioning and Pretesting Effects: a Review of Empirical Research, Theoretical Perspectives, and Implications for Educational

Practice. *Educational Psychology Review*, *35*(4). https://doi.org/10.1007/s10648-023-09814-5

Pi, Z., & Hong, J. (2016). Learning process and learning outcomes of video podcasts including the instructor and PPT slides: A Chinese case. *Innovations in Education and Teaching International*, *53*(2), 135–144. https://doi.org/10.1080/14703297.2015.1060133

Pi, Z., Hong, J., & Yang, J. (2017a). Does instructor's image size in video lectures affect learning outcomes? *Journal of Computer Assisted Learning*, *33*(4), 347–354. https://doi.org/10.1111/jcal.12183

Pi, Z., Hong, J., & Yang, J. (2017b). Effects of the instructor's pointing gestures on learning performance in video lectures. *British Journal of Educational Technology*, *48*(4), 1020–1029. https://doi.org/10.1111/bjet.12471

Pi, Z., Xu, K., Liu, C., & Yang, J. (2020). Instructor presence in video lectures: Eye gaze matters, but not body orientation. *Computers & Education*, *144*, Article 103713. https://doi.org/10.1016/j.compedu.2019.103713

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

Rat für kulturelle Bildung. (2019). *Jugend, Youtube, kulturelle Bildung.* https://www.rat-kulturellebildung.de/fileadmin/user upload/pdf/StudieYouTube Webversionfinal 2.pdf. Accessed March 7, 2021

Reeves, B., & Nass, C. (1996). *The Media Equation: How people treat computers, television, and new media like real people and places*. Cambridge University Press.

Rey, G. D. (2012). A review of research and a meta-analysis of the seductive detail effect. *Educational Research Review*, *7*(3), 216–237. https://doi.org/10.1016/j.edurev.2012.05.003

Rosch, J. L., & Vogel-Walcutt, J. J. (2013). A review of eye-tracking applications as tools for training. *Cognition, Technology & Work*, *15*(3), 313–327. https://doi.org/10.1007/s10111-012-0234-7

Rosenholtz, R. (2016). Capabilities and Limitations of Peripheral Vision. *Annual Review of Vision Science*, *2*, 437–457. https://doi.org/10.1146/annurev-vision-082114-035733

Rosenthal, S., & Walker, Z. (2020). Experiencing Live Composite Video Lectures: Comparisons with Traditional Lectures and Common Video Lecture Methods. *International Journal for the Scholarship of Teaching and Learning*, *14*(1), 1–10. https://doi.org/10.20429/ijsotl.2020.140108

Schneider, S., Beege, M., Nebel, S., Schnaubert, L., & Rey, G. D. (2022). The Cognitive-Affective-Social Theory of Learning in digital Environments (CASTLE). *Educational Psychology Review*, *34*(1), 1–38. https://doi.org/10.1007/s10648-021-09626-5

Schneider, S., Nebel, S., Beege, M., & Rey, G. D. (2020). The retrieval-enhancing effects of decorative pictures as memory cues in multimedia learning videos and subsequent performance tests. *Journal of Educational Psychology*, *112*(6), 1111–1127. https://doi.org/10.1037/edu0000432

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

Schramm, H., & Hartmann, T. (2008). The PSI-Process Scales. A new measure to assess the intensity and breadth of parasocial processes. *Communications*, *33*(4), 385–401. https://doi.org/10.1515/COMM.2008.025

Schroeder, N. L. (2016). A Preliminary Investigation of the Influences of Refutation Text and Instructional Design. *Technology, Knowledge and Learning*, *21*(3), 325–340. https://doi.org/10.1007/s10758-016-9278-8

Schroeder, N. L., & Cenkci, A. T. (2018). Spatial Contiguity and Spatial Split-Attention Effects in Multimedia Learning Environments: a Meta-Analysis. *Educational Psychology Review*, *30*(3), 679–701. https://doi.org/10.1007/s10648-018-9435-9

Short, J., Williams, E., & Christie, B. (1976). *The social psychology of telecommunications*. Wiley.

Skulmowski, A., & Rey, G. D. (2020). Covid-19 as an accelerator for digitalization at a German university: Establishing hybrid campuses in times of crisis. *Human Behavior and Emerging Technologies*, *2*, 212–216. https://doi.org/10.1002/hbe2.201

Stull, A. T., Fiorella, L., & Mayer, R. E. (2021). The case for embodied instruction: The instructor as a source of attentional and social cues in video lectures. *Journal of Educational Psychology*, *113*(7), 1441–1453. https://doi.org/10.1037/edu0000650

Sundararajan, N., & Adesope, O. (2020). Keep it Coherent: A Meta-Analysis of the Seductive Details Effect. *Educational Psychology Review*, *38*(2), 149. https://doi.org/10.1007/s10648-020-09522-4

Sweller, J. (1999). *Instrucional design in technical areas*. ACER.

Sweller, J., Ayres, P., & Kalyuga, S. (2011). *Cognitive load theory*. Springer.

Sweller, J., & Chandler, P. (1994). Why some material is difficult to learn. *Cognition and Instruction*, *12*(3), 185–233. https://doi.org/10.1207/s1532690xci1203_1

Sweller, J., van Merrienboer, J. J., & Paas, F. G. C (1998). Cognitive architecture and instructional design. *Educational Psychology Review*, *10*(3), 251–296. https://doi.org/10.1023/A:1022193728205

Sweller, J., van Merriënboer, J. J. G., & Paas, F. (2019). Cognitive Architecture and Instructional Design: 20 Years Later. *Educational Psychology Review*, *31*(2), 261–292. https://doi.org/10.1007/s10648-019-09465-5

ten Hove, P., & van der Meij, H. (2015). Like it or not. What characterizes YouTube's more popular instructional videos? *Technical Communication*, *62*(1), 48–62.

Toftness, A. R., Carpenter, S. K., Geller, J., Lauber, S., Johnson, M., & Armstrong, P. I. (2018). Instructor fluency leads to higher confidence in learning, but not better learning. *Metacognition and Learning*, *13*(1), 1–14. https://doi.org/10.1007/s11409-017-9175-0

van Alten, D. C., Phielix, C., Janssen, J., & Kester, L. (2019). Effects of flipping the classroom on learning outcomes and satisfaction: A meta-analysis. *Educational Research Review*, *28*, Article 100281. https://doi.org/10.1016/j.edurev.2019.05.003

van Gog, T. (2022). The Signaling (or Cueing) Principle in Multimedia Learning. In R. E. Mayer & L. Fiorella (Eds.), *The Cambridge handbook of multimedia learning* (3rd ed., pp. 221–230). Cambridge University Press. https://doi.org/10.1017/9781108894333.022

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

van Wermeskerken, M., Ravensbergen, S., & van Gog, T. (2018). Effects of instructor presence in video modeling examples on attention and learning. *Computers in Human Behavior*, *89*, 430–438. https://doi.org/10.1016/j.chb.2017.11.038

Võ, M. L.-H., Smith, T. J., Mital, P. K., & Henderson, J. M. (2012). Do the eyes really have it? Dynamic allocation of attention when viewing moving faces. *Journal of Vision*, *12*(13). https://doi.org/10.1167/12.13.3

Wang, H., Pi, Z., & Hu, W. (2019). The instructor's gaze guidance in video lectures improves learning. *Journal of Computer Assisted Learning*, *35*(1), 42–50. https://doi.org/10.1111/jcal.12309

Wang, J., & Antonenko, P. (2017). Instructor presence in instructional video: Effects on visual attention, recall, and perceived learning. *Computers in Human Behavior*, *71*, 79–89. https://doi.org/10.1016/j.chb.2017.01.049

Wang, J., Antonenko, P., & Dawson, K. (2020). Does visual attention to the instructor in online video affect learning and learner perceptions? An eye-tracking analysis. *Computers & Education*, *146*, Article 103779. https://doi.org/10.1016/j.compedu.2019.103779

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

Wilson, K. E., Martinez, M., Mills, C., D'Mello, S., Smilek, D., & Risko, E. F. (2018). Instructor presence effect: Liking does not always lead to learning. *Computers & Education*, *122*, 205–220. https://doi.org/10.1016/j.compedu.2018.03.011

Yuan, M., Zeng, J., Wang, A., & Shang, J. (2021). Would It Be Better if Instructors Technically Adjust Their Image or Voice in Online Courses? Impact of the Way of Instructor Presence on Online Learning. *Frontiers in Psychology*, *12*, Article 746857. https://doi.org/10.3389/fpsyg.2021.746857

Zhang, H., Miller, K. F., Sun, X., & Cortina, K. S. (2020). Wandering Eyes: Eye Movements During Mind- wandering in Video Lectures. *Applied Cognitive Psychology*, *34*(2), 449-464. https://doi.org/10.1002/acp.3632

Zhang, Y., Xu, K., Pi, Z., & Yang, J. (2022). Instructor's position affects learning from video lectures in Chinese context: an eye-tracking study. *Behaviour & Information Technology*, *41*(9), 1988–1997. https://doi.org/10.1080/0144929X.2021.1910731

Zhang, Y., & Yang, J. (2022). Exploring Gender Differences in the Instructor Presence Effect in Video Lectures: An Eye-Tracking Study. *Brain Sciences*, *12*(7). https://doi.org/10.3390/brainsci12070946

# Appendix A

Manuscript 1

Sondermann, C. & Merkt, M. (2023). What is the effect of talking heads in educational videos with different types of narrated slides? *Contemporary Educational Psychology*, *74*, Article 102207. https://doi.org/10.1016/j.cedpsych.2023.102207

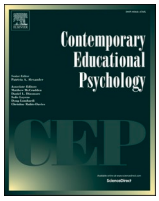
| **Author** | **Author position** | **Scientific ideas %** | **Data generation %** | **Analysis and interpretation %** | **Paper writing %** |
|---|---|---|---|---|---|
| Christina Sondermann | 1 | 75 | 100 | 90 | 90 |
| Martin Merkt | 2 | 25 | 0 | 10 | 10 |

Contemporary Educational Psychology 74 (2023) 102207

Contents lists available at ScienceDirect

Contemporary Educational Psychology

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



# What is the effect of talking heads in educational videos with different types of narrated slides?

Christina Sondermann *, Martin Merkt

*German Institute for Adult Education – Leibniz Centre for Lifelong Learning, Heinemannstraße 12-14, 53175 Bonn, Germany*

A R T I C L E I N F O

*Keywords:*
Talking head
Instructor presence
Video-based learning
Instructional design
Multimedia learning

A B S T R A C T

Video-based learning plays an increasingly important role and thus the optimal design of video-based learning materials attracts the attention of scientists and practitioners alike. In this context, producers of educational videos often include a talking head in their videos, although theory (e.g., cognitive theory of multimedia learning) also suggests potential disadvantages for this format. Since talking heads attract a lot of visual attention, further empirical research is necessary to investigate whether a talking head can hinder learning, especially presented next to graphic-based learning content. To address this research gap, we conducted two online experiments to investigate the effects of a talking head in educational videos with narrated slides (short slideshow lectures) on learning outcomes (i.e., factual knowledge acquisition) and participants' subjective ratings of the learning material (e.g., perceived learning). In Experiment 1 ($N = 96$), we varied whether the instructor's talking head was present or absent in the videos as a between-factor, and whether the visual content on the slides was graphic-based (pictures, diagrams, maps) or text-based (bullet points) as a within-factor (slide type). In Experiment 2 ($N = 184$), we additionally varied as a between-factor whether the contents appeared sequentially or statically all at once (presentation type). Our results showed that the talking head did not affect learning outcomes, regardless of slide type and presentation type of the videos suggesting that the inclusion of a talking head offers neither clear advantages nor disadvantages. Potential explanations for the findings and directions for future research are discussed.

## 1. Introduction

### 1.1. Objective and rationale

When designing learning videos for formal or informal educational contexts, the question arises whether or not to include a talking head in the videos. In this format, a video of a human instructor's head - recorded lecturing into the camera (Hansch et al., 2015) without further didactic functions (e.g., gesturing, pointing, drawing) - is displayed next to the actual learning materials. If we try to answer this question based on theory, this format offers more potential disadvantages than advantages, as a talking head provides few social cues and poses the risk of distraction. Previous research on instructor presence in general (Alemdag, 2022; Henderson & Schroeder, 2021) and their didactic functions (e.g., Beege et al., 2020; Fiorella et al., 2019; Ouwehand et al., 2015; Pi et al., 2017; Stull et al., 2021) has already provided many insightful findings. However, broad evidence on the effects of a human talking head as well as a systematic investigation of potential boundary conditions is still lacking, although this format is frequently used (e.g., in online lectures). Nevertheless, there is a particular need for research on this format, because the talking head attracts a lot of visual attention (Kizilcec et al., 2014) and can thus be potentially distracting, while at the same time its positive effects may be limited because a talking head is characterised by a low level of embodiment (Mayer, 2021). Therefore, the question arises whether there are boundary conditions for the use of a talking head, i.e., whether the talking head has detrimental effects with materials that themselves require a lot of visual attention.

For our investigation, we vary whether a talking head is visible in one corner of the videos next to the learning materials (i.e., short narrated slideshows). By additionally varying the slide type of the learning videos, we aim to investigate whether the talking head has different effects with information which is only displayed visually (in our experiments operationalised with graphic slides) than with information which is also included in the narration (in our experiments operationalised with text slides).

* Corresponding author.
*E-mail addresses:* sondermann@die-bonn.de (C. Sondermann), merkt@die-bonn.de (M. Merkt).

https://doi.org/10.1016/j.cedpsych.2023.102207

Available online 13 July 2023

### 1.2. Theoretical background

Concerning the theoretical foundations about the effects of talking heads in educational videos, different approaches have to be taken into account. First, the social agency theory (SAT, Mayer, 2014) highlights the significant role of social cues in multimedia learning, given that they may activate social responses which in turn motivate learners to engage in more active processing (i.e., enhanced germane cognitive load) and result in enhanced transfer learning performance (also see Cognitive-Affective-Social Theory of Learning in digital Environments, Schneider et al., 2022). In the framework of SAT, the social response is considered an important component which is accompanied by a feeling of social presence (Mayer, 2014). Social presence is defined as the "degree of salience of the other person in the interaction" (Short et al., 1976, p. 65). Further, the embodiment principle of the cognitive theory of multimedia learning (CTML, Mayer, 2021) states that an on-screen instructor's impact on learning is determined by the level of embodiment, e.g., the extent to which the instructor provides human-like gestures or body movements. In this regard, the embodiment principle distinguishes between high and low embodiment, predicting learning-enhancing effects for instructors with a high level of embodiment. In contrast, a talking head who does not show any gestures or body movements is characterised by a low level of embodiment. Hence, the beneficial effects of instructor presence as postulated by the SAT should only apply to a limited extent or not at all to this format. Correspondingly, the image principle of multimedia learning (Mayer, 2021) stresses that learners do not learn better when the instructor's static image is presented on the screen.

Second, according to cognitive load theory (CLT, Sweller, 1999; Sweller et al., 2011) and CTML (Mayer, 2021), the working memory capacity is limited (Baddeley, 1992). The CLT distinguishes between three different types of cognitive load, with intrinsic load resulting from the difficulty of the task, whereas extraneous load stems from the instructional design of the learning material (Klepsch et al., 2017; Sweller et al., 1998). This extraneous load should be reduced in order to preserve the remaining capacity for active cognitive processing, which is relevant for learning and is referred to as the germane cognitive load (Sweller et al., 1998). Based on this, the coherence principle of multimedia learning suggests that "people learn more deeply from a multimedia message, when extraneous material is excluded rather than included" (Mayer & Fiorella, 2014, p. 279). Thus, when a design element such as the talking head does not contribute to the learning task, it should be excluded in order to reduce learners' extraneous cognitive load.

Further, a talking head can lead to visual competition between the instructor's face and the learning content in the videos resulting in split attention (Wang & Antonenko, 2017). Due to the need of mentally integrating the multiple sources of information, split attention leads to increased extraneous cognitive load (Ayres & Sweller, 2014). Hence, the split-attention principle suggests to "avoid materials that require learners to split their attention" (Ayres & Sweller, 2014, p. 206). When a substantial part of relevant information in the educational video is delivered by the visual content (e.g., graphics, diagrams, maps), a talking head might distract learners from this relevant content and thus hinder learning.

Taken together, in the light of the reported theoretical background, research on talking heads is highly relevant, since the potential advantages of instructor presence only play a very limited role in this format due to its low embodiment, while at the same time its potential disadvantages (e.g., split attention, distraction, cognitive overload) should come into play (e.g., Mayer, 2021). Against the background of theory (e.g., split-attention principle), detrimental effects of a talking head on learning are therefore conceivable, especially with graphic-based learning content.

### 1.3. Literature review

Since research on the effects of talking heads in educational videos is still rather limited (Kizilcec et al., 2014; Kizilcec et al., 2015; Kokoç et al., 2020; Ng & Przybylek, 2021; Rosenthal & Walker, 2020; Yuan et al., 2021), we present the research on human instructor presence in learning videos in general. Hence, we also illuminate the findings in light of the instructor's respective level of embodiment, whenever this information was available from the particular studies. In the context of instructor presence, research on video modeling examples (VME) also provides many insightful findings, as they involve videos in which a model demonstrates and/or explains a solution procedure (van Gog & Rummel, 2010; van Gog et al., 2014). Many studies on VME investigate how existing models should ideally be presented or what they should do to improve learning (e.g., model-observer similarity, Hoogerheide et al., 2018; level of embodiment, Fiorella & Mayer, 2016; van Gog et al., 2014; gaze cues, Ouwehand et al., 2015). Whereas these studies offer important insights into the design of VME, our experiments focus on whether it is useful to include an instructor at all if no pedagogically compelling reasons arise from the requirements of the materials (e.g., demonstrating a manual procedure). Therefore, we focus our literature review on studies on VME that also examine an absent instructor condition (Fiorella et al., 2019; Hoogerheide et al. 2014; van Wermeskerken et al., 2018).

#### 1.3.1. Learning outcomes

In line with SAT, when onscreen instructors display a high level of embodiment, they tend to aid learning (e.g., Yu, 2021); for instance, regarding recall performance for easy videos (Wang & Antonenko, 2017), transfer performance (Wang, Antonenko, Keil, & Dawson, 2020), or the retention of spoken explanations (Colliot & Jamet, 2018). Consistent with the embodiment principle (Mayer, 2021), learners in the respective studies could see the instructors' whole upper body and gestures.

In contrast, some studies report no learning differences between conditions with and without an instructor neither for retention nor for transfer knowledge (Fiorella et al., 2019; Homer et al., 2008; Hoogerheide et al., 2014; Moreno et al., 2001; Ng & Przybylek, 2021; van Wermeskerken et al., 2018), and neither for immediate nor for delayed recall (Kizilcec et al., 2014). The instructors' embodiment in these studies was in some cases rather high (van Wermeskerken et al., 2018), and in some cases rather low due to the use of a talking head (Kizilcec et al., 2014; Ng & Przybylek, 2021; Rosenthal & Walker, 2020).

In line with CLT and CTML, when onscreen instructors display a low level of embodiment, they do not offer much improvement in learning (Kizilcec et al., 2014; Ng & Przybylek, 2021; Rosenthal & Walker, 2020; but see Kokoç et al., 2020). Two studies even suggest learning-impeding effects of visible instructors (Wilson et al., 2018, Experiment 5) and talking heads in particular (Yuan et al., 2021), with additional conditions including a face-beautified instructor and a virtual instructor also leading to learning losses in the latter study. Fittingly, in previous studies using eye-tracking (Colliot & Jamet, 2018; Pi & Hong, 2016; Stull et al., 2018, 2021; van Wermeskerken et al., 2018; Wang, Antonenko, & Dawson, 2020; Wang & Antonenko, 2017), the instructors attracted a considerable amount of visual attention, also including instructors with a low level of embodiment, i.e. talking heads (Kizilcec et al., 2014). This also resulted in less attention being paid to the learning content (Colliot & Jamet, 2018; van Wermeskerken et al., 2018; Wang, Antonenko, & Dawson, 2020). Accordingly, Wilson et al. (2018) argue that distraction by an instructor may be more harmful with visual video content than with content which is covered by narration, without testing this explanation.

#### 1.3.2. Subjective ratings

In addition to objective learning outcomes, learners' subjective ratings of the learning materials are also of central importance, especially

in non-supervised learning, because learners' subjective experience of learning may contribute to the decision to watch or continue watching certain videos (Guo et al., 2014). Additionally, since ratings of perceived learning do not always correspond to those of actual learning (e.g., Toftness et al., 2018), some empirical studies also consider this subjective measure of learning. For example, prior research showed that both talking heads examined in an observational field study (Kizilcec et al., 2015) and visible instructors providing gestures (Wang & Antonenko, 2017) enhanced learners' ratings of perceived learning compared to videos without instructor (Wilson et al., 2018).

Since SAT particularly emphasises social processes by assuming that social cues can result in social responses and learning-enhancing effects (Mayer, 2014), some studies also examined effects of instructor presence on perceived social presence (Colliot & Jamet, 2018; Homer et al., 2008), with two studies also investigating talking heads (Ng & Przybylek, 2021; Yuan et al., 2021). Although the instructor was characterized by a comparatively high level of embodiment in at least one of the studies (Colliot & Jamet, 2018), these studies did not show any effects of the instructor on social presence ratings.

While CLT and CTML (cf. coherence principle, split-attention principle) would suggest an increase in extraneous cognitive load due to instructor presence, only a few studies revealed that a visible instructor enhanced different indicators of cognitive load (Homer et al., 2008; Hong et al., 2018). Other studies showed no effects of instructor presence on extraneous cognitive load or mental effort, examining both a talking head (Ng & Przybylek, 2021) and an instructor with visible upper body and gestures (Colliot & Jamet, 2018). In contrast, Wang, Antonenko, Keil, and Dawson (2020) even found lower extraneous and intrinsic cognitive load ratings when an instructor showing gestures was present.

#### 1.3.3. Overview of talking head effects

Taken together, with respect to subjective ratings, talking heads had positive effects on learners' perceived learning (Kizilcec et al., 2015), whereas they did not affect ratings of social presence (Ng & Przybylek, 2021, Yuan et al., 2021) and cognitive load (Ng & Przybylek, 2021). In terms of learning outcomes, consistent with the embodiment principle (Mayer, 2021), learning-enhancing effects of an instructor mostly occurred when the instructor showed a high level of embodiment, while instructors with a low level of embodiment such as talking heads often had no effect on learning (Kizilcec et al., 2014; Ng & Przybylek, 2021; Rosenthal & Walker, 2020; but see Kokoç et al., 2020). One study examining a Chinese sample even suggests that a talking head may hinder learning (Yuan et al., 2021). Despite their low level of embodiment, talking heads also attracted a lot of visual attention (Kizilcec et al., 2014). However, since previous research did not systematically vary the nature of the slide contents next to the instructor, precise information on how the contents were visually presented (i.e., whether they were graphic-based or text-based), is often missing. Hence, the question remained open whether talking heads could be detrimental to learning when presented next to graphic-based content that is not covered by the spoken narration making it worthwhile to systematically vary the content presented next to the talking head.

### 1.4. The present research

In summary, theory mainly suggests potential disadvantages (distraction, split attention) for the talking head format (low embodiment), even though this is not consistently evident in previous research. However, since talking heads attract a lot of visual attention (Kizilcec et al., 2014), they might be detrimental to learning, especially with graphic-based content (cf. Wilson et al., 2018). In this regard, the findings of Colliot and Jamet (2018) indicate that instructor presence should be less disturbing for information that is also presented in spoken narration than for information that is only presented visually. Therefore, extending research suggesting that the effects of instructor presence may be moderated by the features of the learning materials (e.g., Hong et al., 2018; Wang & Antonenko, 2017), we systematically vary the slide type of the videos with narrated slides (graphic vs. text) as a potential boundary condition. In doing so, we manipulate whether the relevant content appears only visually (i.e., graphic slides containing pictures, diagrams, maps) or appears visually but is also covered by the narration (i.e., text slides containing bullet points).

## 2. Experiment 1

### 2.1. Hypotheses

Based on theory and previous research outlined above, we preregistered hypotheses and corresponding analyses on Open Science Framework (OSF, https://osf.io/vcgp5). First, Experiment 1 addressed whether the slide type in the videos moderates the talking head effects on learning. In this regard, we expected an interaction of slide type and talking head (Hypothesis 1). Since the talking head attracted a lot of visual attention in prior research, its presence might have a detrimental effect on learning outcomes (i.e., acquisition of factual knowledge) for videos with graphic slides, where the visual content is relevant for learning. For videos with text slides, where the relevant content is also covered by the narration, we expected no effects of the talking head on learning outcomes.

Second, based on prior research (e.g., Kizilcec et al., 2015), we expected perceived learning to be higher for videos with talking head than for those without (Hypothesis 2).

In order to identify explanatory mechanisms for potential effects of the talking head on objective learning outcomes, we recorded ratings on social presence and cognitive load (i.e., intrinsic, germane, and extraneous cognitive load) based on theory (CLT, CTML, SAT). We did not derive hypotheses for these variables due to inconsistent or missing empirical findings, but instead preregistered exploratory analyses.

### 2.2. Method

#### 2.2.1. Participants

Since previous research reports heterogeneous findings, we decided to conduct an a priori power analysis to determine a reasonably large sample size which would allow us to reliably detect medium (Cohen, 1988) desirable (Hattie, 2008) effect sizes with justifiable resources. We conducted the power analysis with G*Power 3.1.9.4 (Faul et al., 2007) which suggested to collect at least $N = 90$ participants ($\alpha = 0.05$, $1\text{-}\beta = 0.80$, $d = 0.6$) to find medium effects. To achieve equal balancing, we preregistered to fill up the respective groups to the size of the largest group in case randomisation would result in an unequal distribution of participants across groups. A total of $N = 104$ German-speaking participants recruited via the online platform Prolific completed the online experiment. Based on preregistered exclusion criteria, we excluded eight participants because they had not seen all videos ($n = 1$), they took notes on the videos ($n = 5$), their videos had been interrupted ($n = 1$), or they did not provide any information on their native language ($n = 1$). This resulted in a final sample size of $N = 96$ (26 female, $M_{age} = 29.01$ years, $SD_{age} = 8.64$).

#### 2.2.2. Learning materials

The learning materials consisted of six videos with narrated slides providing factual information about six different topics in German language. Three videos contained only graphic slides, three videos only text slides. In total, all graphic-based videos together lasted 11 min and 50 s (varying between 143 and 314 s), all text-based videos 11 min and 15 s (varying between 159 and 304 s). For both graphic- and text-based videos, one slideshow consisted of 6 slides, one of 7 slides, and one of 9 slides. We chose video topics for which we assumed that participants had little or no prior knowledge: The three graphic-based videos dealt with the archipelago of South Georgia and the South Sandwich Islands

(e.g., geography, location), the art epochs between 1400 and 1900 (e.g., artists and works), and the Triangular Model as a statistical representation (e.g., use, structure, advantages, and disadvantages). Contents of the graphic slides were presented sequentially according to the narration and were partially highlighted by temporary framing them. In graphic-based videos, the narration referred to the graphic contents, but the visual contents were not identical and rather complementary to the narration containing information which did not appear in the narration.

The three text-based videos provided information about the African country Burkina Faso (e.g., location, neighbouring countries, inhabitants), the biography of Rudolf Höß (e.g., childhood, career, death), and the songbird blackcap (e.g., appearance, distribution, diet). Within the text slides, the contents appeared in bullet points as soon as the narration referred to them. Thus, there was a large overlap between the narration and the written bullet points allowing learners to obtain all relevant information also by listening to the audio narration.

In the talking head condition, a video of the instructor's talking head was visible in either the lower left or right corner (balanced) of the video and did not overlap with the learning content on the slides. The instructor showed no gestures and simply spoke into the camera representing a low level of embodiment (see Fig. 1). Learners in all conditions heard the same narration.

### 2.2.3. Measures

*2.2.3.1. Prior knowledge.* Learners indicated their prior knowledge about the six learning contents on 7-point Likert scales ranging from 1 (*no prior knowledge*) to 7 (*a lot of prior knowledge*) with one question per topic (e.g., "How much do you know about the African country Burkina Faso?", all questions translated for the manuscript). Each participant's ratings were averaged to prior knowledge scores for graphic- and text-based topics. Cronbach's alpha was $\alpha = 0.60$ for graphic-based topics and $\alpha = 0.61$ for text-based topics.

*2.2.3.2. Knowledge test / learning outcomes.* After participants watched all videos, we assessed their learning outcomes using a self-designed knowledge test consisting of 8 multiple-choice recall questions per topic (i.e., 48 questions) that referred to facts explicitly taught in the videos. We decided to capture factual knowledge, because learners are most likely to miss facts if they are distracted by the talking head. Each question consisted of four alternative answers (from which one was correct) and a no-answer option. For graphic-based topics, some questions included pictures as question or answer options. The time limit for each question was 60 s. We assigned one point for each correctly answered question, so that participants could achieve a maximum of 24 points for graphic-based videos and 24 for text-based videos. Cronbach's alpha was $\alpha = 0.73$ for the graphic-based topics and $\alpha = 0.62$ for the text-based topics.

*2.2.3.3. Ratings of the learning material*

*2.2.3.3.1. Perceived learning.* We assessed perceived learning both directly after each video and as an overall rating after the entire learning phase. Since both ratings resulted in the same interpretation, we only report data and analyses on the video-specific rating in the Results sections. Learners rated their perceived learning ("How much have you learned from the video(s)?"), adapted from Kizilcec et al. (2015) and Wang and Antonenko (2017) on a 7-point Likert scale from 1 (*nothing at all*) to 7 (*a lot*). Mean scores were calculated for videos with graphic ($\alpha = 0.52$) and text slides ($\alpha = 0.80$).

*2.2.3.3.2. Social presence.* The perceived social presence was assessed with three adapted items on a 7-point Likert scale from 1 (*strongly disagree*) to 7 (*strongly agree*). Two items ("I felt that the instructor was present", "I felt like the instructor was in the same room as me") were taken from a questionnaire used by Kizilcec et al. (2015). The third ("I felt like the instructor addressed me personally") was one item from the PSI process scales (Schramm & Hartmann, 2008) already applied by Beege et al. (2017). For each participant, ratings for the three items were averaged to an overall social presence score. Cronbach's alpha was $\alpha = 0.87$.

*2.2.3.4. Cognitive load*

*2.2.3.4.1. **Intrinsic, germane, and extraneous cognitive load.*** To assess cognitive load, we used the scale of Klepsch et al. (2017). Learners had to answer eight items on 7-point Likert scales from 1 (*absolutely wrong*) to 7 (*absolutely right*). Two items addressed intrinsic cognitive load and three items each addressed germane and extraneous cognitive load. Reliabilities were $\rho = .74$[1] for intrinsic, $\alpha = 0.65$ for germane, and $\alpha = 0.80$ for extraneous cognitive load.

### 2.2.4. Design

The experiment followed a 2x2-factorial design including the between-subjects factor talking head (present vs. absent) and the within-subjects factor slide type (videos consisting of graphic vs. text slides). Thus, whereas participants were randomly assigned to one of the two talking head conditions, each participant watched three videos with graphic and three videos with text slides. In the talking head condition, we balanced whether the talking head appeared in the lower right or left corner.

### 2.2.5. Procedure

The online experiment was conducted with LimeSurvey. Throughout the experiment, we collected different variables, which are not presented in this manuscript due to space limitations. For completeness, we mention these variables when describing the procedure. All relevant information and results for these variables can be found on OSF (https://osf.io/xug9c/?view_only=7b149959778c40c3b9fe51fb988f823f). After giving informed consent, participants indicated their prior knowledge and their interest about the contents of the educational videos. After ensuring that audio and video worked properly, all participants watched the six videos in the same fixed order, alternating between text- and graphic-based contents. The videos started automatically and did not allow for user interactions. After each video, we assessed invested mental effort, perceived difficulty, and perceived learning. After watching all the videos, learners gave a detailed assessment of cognitive load. Subsequently, they had to rate the learning materials (i.e., perceived learning, satisfaction, professionalism, joy of learning, and social presence). Due to time constraints, we did not record all of these items (e.g., eight items in the cognitive load scale) after each of the six videos because our design allows us to draw conclusions about talking head effects (between-subjects factor) using the measurement across different videos; whereas the assessment of some of the items after each video (e.g., perceived learning) allows for analyses regarding the interplay of talking heads and slide type (within-subjects factor). Further, we recorded learners' verbal-visual learning style rating (Mayer & Massa, 2003), which explicitly asks for participants' learning preference[2]. This was followed by the knowledge test. Further, participants answered demographic questions and control questions designed to exclude participants based on preregistered exclusion criteria (e.g., use of a search engine, problems with video playback). Finally, participants were debriefed. The procedure was approved by the local ethics committee.

[1] For scales with only two items, we used the Spearman-Brown coefficient (ρ) instead of Cronbach's alpha, because the former is more appropriate for determining reliability of two items (Eisinga et al., 2013).

[2] We included this measure because some studies suggested an interplay between instructor presence and learning preference (Homer et al., 2008; Kizilcec et al., 2015). Therefore, we intended to rule out prior differences among our between-subjects conditions, which we were able to do (see Online Appendix).

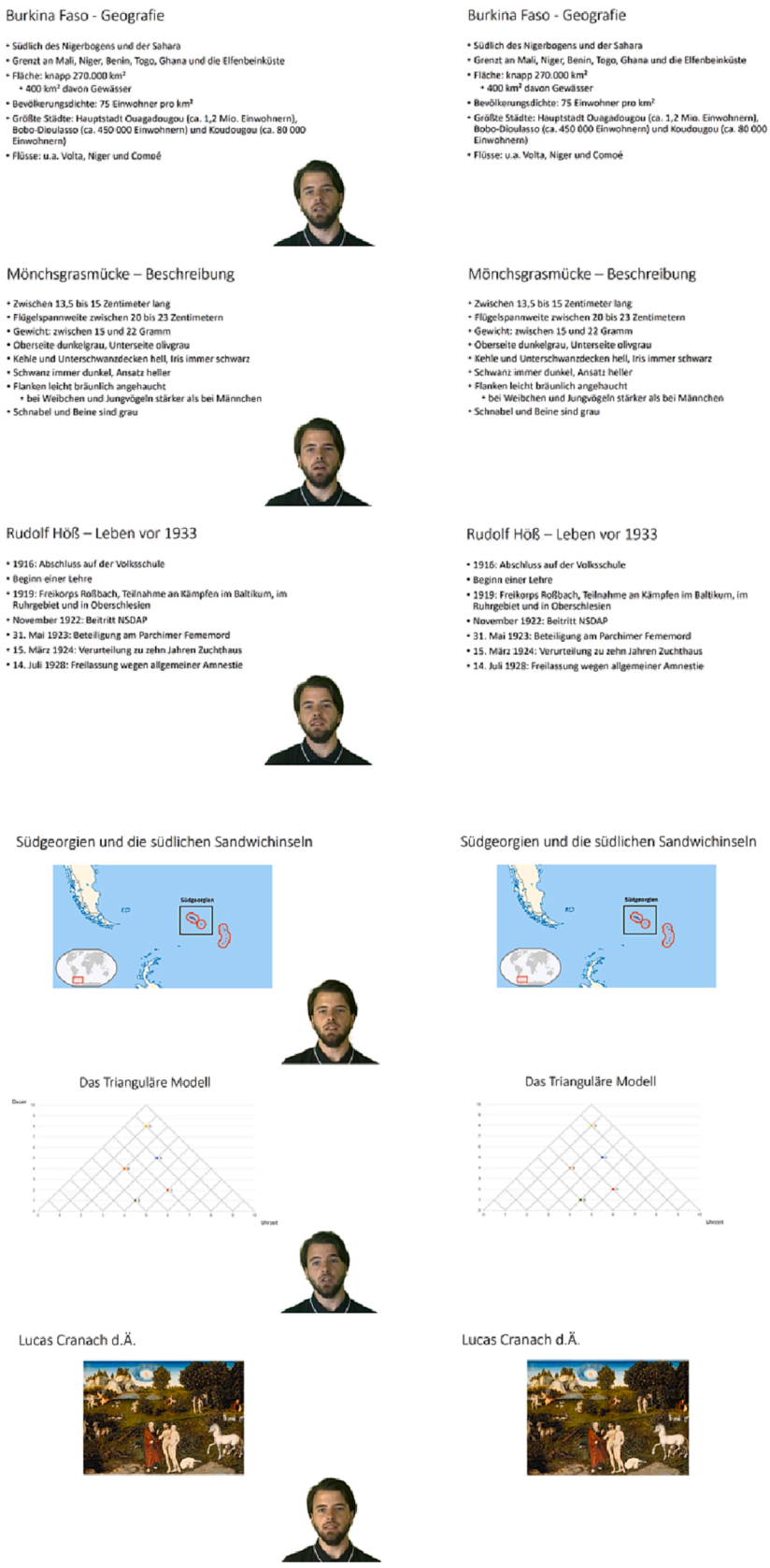



**Fig. 1.** Screenshots of text-based (Rows 1 to 3) and graphic-based (Rows 4 to 6) videos of Experiment 1 with talking head (left) and without talking head (right). Picture credits: All pictures except the pictures in Rows 4 and 6 were self-generated. Row 4: TUBS, South Georgia and the South Sandwich Islands in its region. This graphic contains elements which have been taken or were adapted from the following graphic (CC BY-SA 3.0): via https://commons.wikimedia.org/w/index.php?curid = 15431607, Row 6: Lucas Cranach the Elder, Adam and Eve-Paradise (Public domain): via https://commons.wikimedia.org/wiki/File:Lucas_Cranach_(I)_-_Adam_and_Eve-Paradise_-_Kunsthistorisches_Museum.jpg.

#### 2.2.6. Analytical strategy

Regarding the dependent variables learning outcomes and perceived learning, we applied 2x2-factorial ANOVAs with the independent variables talking head (between) and slide type (within). For completeness, potential main effects of slide type are reported in the Results section, but not interpreted due to different video contents (see Section 4.3). For social presence and cognitive load, we applied one-way ANOVAs with talking head as between-subjects factor because these general ratings after the learning phase did not allow for a differentiation between graphic-based and text-based videos. Scripts and data can be downloaded at OSF (https://osf.io/xug9c/?view_only=7b149959778c40c3b9fe51fb988f823f).

### 2.3. Results and discussion

Descriptive data on the key measures (learning outcomes, perceived learning, social presence, and cognitive load) can be found in Table 1. In the following, we report results concerning these main research questions. Descriptive data for further variables and analyses regarding these variables (e.g., perceived difficulty, satisfaction, and professionalism) can be found at OSF.

#### 2.3.1. Did the groups differ on individual prerequisites?

We conducted analyses on individual prerequisites to ensure that participants did not differ in terms of age, gender, and self-reported prior knowledge on the covered topics. For age, the ANOVA revealed no pre-existing differences between conditions with and without talking head, $F(1, 94) = 2.51$, $p = .117$, $\eta_p^2 = 0.026$. Regarding gender, the chi-square test also indicated no differences between talking head conditions, $\chi^2(1) = 1.90$, $p = .168$, $\varphi = 0.141$. For prior knowledge, the ANOVA revealed a significant main effect of slide type with higher self-reported prior knowledge levels for the topics of the graphic-based videos ($M = 1.46$, $SD = 0.61$) compared to text-based videos ($M = 1.30$, $SD = 0.49$), $F(1, 94) = 8.44$, $p = .005$, $\eta_p^2 = 0.082$. There was no significant main effect of talking head and no significant interaction, both $F < 1.28$, both $p > .262$.

#### 2.3.2. Preregistered analyses

*2.3.2.1. Did the talking head negatively affect learning outcomes for graphic-based materials?.* Contrary to Hypothesis 1, we found no interaction Talking Head × Slide Type, $F < 1$. Further, there was no main effect of talking head, $F < 1$, but a main effect of slide type, $F(1, 94) = 5.65$, $p = .019$, $\eta_p^2 = 0.057$, with higher scores for text-based videos ($M = 14.84$, $SD = 3.59$) than for graphic-based videos ($M = 13.84$, $SD = 4.14$). Thus, regardless of slide type, the talking head had no effects on learning outcomes contrary to Hypothesis 1. The absence of a talking head effect on objective learning outcomes is in line with some previous studies (Homer et al., 2008; Kizilcec et al., 2014; van Wermeskerken et al., 2018). However, the missing interaction with slide type contradicts Hypothesis 1 which we derived based on CTML and CLT as well as studies showing that instructors' faces attract learners' visual attention (Colliot & Jamet, 2018; Kizilcec et al., 2014). A potential reason for the missing interaction might lie in the design of our learning materials: The slides in our videos (both graphic and text slides) were animated so that the contents appeared sequentially, meaning that the bullet points in the text slides and the visual elements in the graphic slides appeared one by one. Since the onset of movements attracts attention (Abrams & Christ, 2003), it is conceivable that this sequential presentation repeatedly directed learners' attention away from the talking head back to the relevant contents on the slides (also see Moon & Ryu, 2021). This assumption would be in line with the signaling principle of multimedia learning (Mayer & Fiorella, 2014; van Gog, 2014), which states that visual cues that direct learners' attention to relevant information can result in better learning. Accordingly, previous research showed that a step-by-step presentation of graphic-based contents can have a beneficial effect on learning (Fiorella et al., 2019; Jamet et al., 2008).

*2.3.2.2. Did the talking head positively affect ratings of perceived learning?.* For perceived learning, there was no main effect of talking head, $F < 1$. Further, the interaction Talking Head × Slide Type was not significant, $F < 1$. However, there was a main effect of slide type, $F(1, 94) = 15.78$, $p < .001$, $\eta_p^2 = 0.144$. Values were higher for text-based videos ($M = 4.04$, $SD = 1.23$) than for graphic-based videos ($M = 3.65$, $SD = 1.10$). The missing effect of the talking head on perceived learning is not consistent with Hypothesis 2 and the findings of prior studies suggesting positive effects of instructors (including talking heads) on perceived learning (e.g., Kizilcec et al., 2015; Wang & Antonenko, 2017). This could be explained by the fact that in studies observing positive effects, learners had a direct comparison between videos with and without instructor, for example by watching both videos with and without instructor, or by choosing between both formats.

#### 2.3.3. Exploratory analyses

Regarding social presence ($M = 2.60$, $SD = 1.41$), there was no significant effect of talking head, $F < 1$. This missing effect of the talking head is consistent with previous research (e.g., Ng & Przybylek, 2021; Yuan et al., 2021). Further, there were no effects of talking head on intrinsic ($M = 4.49$, $SD = 1.49$), germane ($M = 5.09$, $SD = 1.02$), and extraneous cognitive load ($M = 4.63$, $SD = 1.32$), all $F < 1$. Taken together, exploratory analyses revealed no impact of the talking head on learners' ratings.

## 3. Experiment 2

Considering the findings of Experiment 1, it is possible that an attention-guiding function of the sequential presentation compensated for the potentially distracting effect of the talking head. Therefore, we conducted a second experiment in which we additionally varied the presentation type of the slide content (sequential vs. static) as between-subjects factor in addition to the factors investigated in Experiment 1. In the static presentation, the contents of a slide were shown all at once

**Table 1**
Means (and Standard Deviations in Parentheses) for Experiment 1.

| Instructor | Slide type | Knowledge test | Perceived learning | Prior knowledge | Social presence | Intrinsic cognitive load | Germane cognitive load | Extraneous cognitive load |
|---|---|---|---|---|---|---|---|---|
| Present | Graphic | 14.29 (3.89) | 3.67 (1.03) | 1.42 (0.48) | | | | |
| ($n = 48$) | Text | 14.88 (3.87) | 4.06 (1.29) | 1.24 (0.36) | | | | |
| | Overall | 14.58 (3.46) | 3.86 (1.07) | 1.33 (0.34) | 2.57 (1.33) | 4.54 (1.47) | 5.04 (1.01) | 4.65 (1.35) |
| Absent | Graphic | 13.40 (4.36) | 3.63 (1.17) | 1.51 (0.72) | | | | |
| ($n = 48$) | Text | 14.81 (3.32) | 4.02 (1.17) | 1.37 (0.59) | | | | |
| | Overall | 14.10 (3.10) | 3.82 (1.05) | 1.44 (0.59) | 2.64 (1.50) | 4.44 (1.52) | 5.14 (1.04) | 4.61 (1.31) |
| Overall | Graphic | 13.84 (4.14) | 3.65 (1.10) | 1.46 (0.61) | | | | |
| ($N = 96$) | Text | 14.84 (3.59) | 4.04 (1.23) | 1.30 (0.49) | | | | |
| | Overall | 14.34 (3.28) | 3.84 (1.06) | 1.38 (0.48) | 2.60 (1.41) | 4.49 (1.49) | 5.09 (1.02) | 4.63 (1.32) |

*Note.* The left part of the table (video-specific measures) refers to 2x2-factorial ANOVAs, the right part (cross-video measures) refers to the one-way ANOVAs.

instead of fading them in one by one.

### 3.1. Hypotheses

Experiment 2 served the dual purpose of (1) replicating the findings of Experiment 1 while (2) testing our abovementioned explanation of the findings. For learning outcomes, we expected the presentation type (sequential vs. static) to affect the interplay of talking head and slide type which should be reflected in a significant three-way interaction (Hypothesis 1): In the conditions with sequential presentation of the content, we expected no significant interaction between talking head and slide type, replicating the results of Experiment 1. However, for static presentation of the content, there is no movement on the slides to draw learners' attention away from the talking head back to the content. Therefore, we expected a learning-impeding effect of the talking head for graphic-based content, whereas we did not expect a detrimental effect for text-based content.

Further, since previous research using graphic-based learning materials reported positive effects of sequential presentation on learning outcomes compared to a static presentation (Fiorella et al., 2019; Jamet et al., 2008), we expected better learning outcomes for graphic slides with sequential presentation than for graphic slides with static presentation (Hypothesis 2.1). It remains an open research question whether the sequential presentation has also positive effects for text-based videos. Further, for videos with graphic slides, detrimental talking head effects should be apparent for static, but not for sequential presentation (Hypothesis 2.2).

Regarding the subjective rating scales (perceived learning, social presence), we expected to replicate the results of Experiment 1 in which we found no effects of the talking head (Hypothesis 3). As in Experiment 1, we preregistered to explore the effects of our independent variables on cognitive load. Again, hypotheses were preregistered on OSF (https://osf.io/r36wn).

### 3.2. Method

#### 3.2.1. Participants

We preregistered to collect data of $N = 184$ participants. Since we did not observe any effects in Experiment 1, we decided to further decrease the specified effect size in the power analysis for the current experiment to detect medium to small effects according to Cohen (1988) based on considerations regarding a desirable effect size (Hattie, 2008). Again, we had conducted an a priori power analysis with G*Power 3.1.9.4 (Faul et al., 2007) which suggested a required sample size of at least $N = 152$ participants ($\alpha = 0.05$, 1-$\beta = 0.80$, $f = 0.23$, $\eta_p^2 = 0.05$). To account for balancing and to slightly increase power, we preregistered a sample size of $N = 184$ based on additional available resources.

A total of $N = 202$ German-speaking participants completed the online experiment via Prolific. Based on preregistered exclusion criteria, we excluded 18 participants because they claimed to know the instructor ($n = 2$), not all videos were played with sound ($n = 1$), they used search engines to answer the knowledge questions ($n = 1$), they took notes on the videos ($n = 10$), their videos had been interrupted ($n = 1$), they did not see the whole video on their screen ($n = 1$), and they exceeded time limits regarding video viewing time and the knowledge test as specified in the preregistration ($n = 2$). This resulted in the previously determined final sample size of $N = 184$ (77 female, 103 male, 4 diverse, $M_{age} = 28.68$ years, $SD_{age} = 8.47$).

#### 3.2.2. Learning materials

We used four of the six videos from Experiment 1, two videos with graphic slides (i.e., South Georgia, Triangular Model) and two videos with text slides (i.e., Burkina Faso, Rudolf Höß). We dropped one video (i.e., art topic) because it did not allow for static presentation of the visual information. Consequently, we also excluded a video with text slides that took a similar amount of time (i.e., blackcap topic) to maintain the same number of videos in both conditions. We implemented the sequential presentation in the same way as in Experiment 1 and only adapted minor details to allow static presentation for all slides. In the static presentation condition, the slide content (i.e., graphic elements for graphic slides, bullet points for text slides) appeared all at once. In the talking head conditions, we again balanced whether the talking head was visible in the lower left or right corner of the videos.

#### 3.2.3. Measures

*3.2.3.1. Individual prerequisites.* Prior knowledge for the four topics was recorded exactly as in Experiment 1. Reliabilities were $\rho = 0.55$ for graphic-based and $\rho = 0.43$ for text-based topics.

*3.2.3.2. Knowledge test / learning outcomes.* Questions of the knowledge test were the same as in Experiment 1 (except for the questions referring to the two excluded videos). Therefore, the knowledge test consisted of 32 questions (8 questions per video), resulting in a maximum score of 32. Cronbach's alpha was $\alpha = 0.78$ for the 16 questions on graphic-based videos and $\alpha = 0.53$ for the 16 questions on text-based videos.

*3.2.3.3. Ratings of the learning material and cognitive load.* The variables perceived learning ($\rho = 0.37$ for graphic-based, $\rho = 0.70$ for text-based videos), social presence ($\alpha = 0.89$), intrinsic ($\rho = 0.76$), germane ($\alpha = 0.34$), and extraneous cognitive load ($\alpha = 0.84$) were collected identical to Experiment 1.

#### 3.2.4. Design

Experiment 2 followed a 2x2x2-design with talking head (present vs. absent) and presentation type (sequential vs. static presentation) as between-subjects factors and slide type (graphic vs. text) as within-subjects factor. Participants were randomly assigned to one of the four different between-conditions. In all conditions, they saw four educational videos consisting of narrated slides with two videos only containing graphic and two only containing text slides.

#### 3.2.5. Procedure and analytical strategy

The procedure, which included the adapted materials, was identical to Experiment 1. For learning outcomes and perceived learning, we applied 2x2x2-factorial ANOVAs with the independent variables talking head (between), presentation type (between), and slide type (within). For the remaining variables, we applied 2x2-factorial ANOVAs with talking head and presentation type as between-subjects factors. To explain potential 2-way and 3-way interactions, corresponding post-hoc tests were preregistered. Scripts and data can be downloaded at OSF (https://osf.io/xug9c/?view_only=7b149959778c40c3b9fe51fb988f823f).

### 3.3. Results and discussion

Descriptive data for the key variables are provided in Table 2. Again, we focus on our main research questions in the Results section, whereas descriptive data and further analyses regarding variables such as perceived difficulty, satisfaction, and professionalism can be found at OSF.

#### 3.3.1. Did the groups differ on individual prerequisites?

For age, the ANOVA revealed no main effects or interactions, all $F < 2.49$, all $p > .116$, indicating that there were no pre-existing age differences in the between-subjects conditions. Regarding gender, the chi-square test also indicated no differences between the four between-subjects conditions, $\chi^2(6) = 6.92$, $p = .328$, $\varphi = 0.194$. For prior knowledge, the ANOVA revealed higher self-reported prior knowledge scores for text-based videos ($M = 1.40$, $SD = 0.60$) compared to graphic-based videos ($M = 1.21$, $SD = 0.48$), $F(1, 180) = 16.20$, $p < .001$, $\eta_p^2 = 0.083$. There were no other significant main effects and interactions, all

**Table 2**
Means (and Standard Deviations in Parentheses) for Experiment 2.

| Condition | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| Presentation type | Instructor | Slide type | Knowledge test | Perceived learning | Prior knowledge | Social Presence | Intrinsic cognitive load | Germane cognitive load | Extraneous cognitive load |
| Sequential | Present | Graphic | 11.76 (3.27) | 4.33 (1.11) | 1.20 (0.43) | | | | |
| | ($n$ = 46) | Text | 10.07 (2.94) | 4.45 (1.17) | 1.30 (0.50) | | | | |
| | | Overall | 10.91 (2.54) | 4.39 (0.91) | 1.25 (0.40) | 2.99 (1.70) | 4.42 (1.52) | 5.06 (0.91) | 4.20 (1.50) |
| | Absent | Graphic | 11.20 (3.47) | 4.26 (1.21) | 1.20 (0.50) | | | | |
| | ($n$ = 46) | Text | 10.43 (2.45) | 4.29 (1.33) | 1.29 (0.43) | | | | |
| | | Overall | 10.82 (2.43) | 4.28 (1.06) | 1.24 (0.38) | 2.69 (1.50) | 4.38 (1.29) | 5.17 (0.84) | 4.58 (1.23) |
| | Overall | Graphic | 11.48 (3.37) | 4.29 (1.15) | 1.20 (0.46) | | | | |
| | ($n$ = 92) | Text | 10.25 (2.70) | 4.37 (1.25) | 1.30 (0.46) | | | | |
| | | Overall | 10.86 (2.47) | 4.33 (0.99) | 1.25 (0.39) | 2.84 (1.60) | 4.40 (1.40) | 5.11 (0.87) | 4.39 (1.38) |
| Static | Present | Graphic | 10.85 (3.50) | 3.90 (1.25) | 1.24 (0.49) | | | | |
| | ($n$ = 46) | Text | 10.22 (2.48) | 4.37 (1.25) | 1.41 (0.67) | | | | |
| | | Overall | 10.53 (2.60) | 4.14 (1.13) | 1.33 (0.45) | 2.41 (1.37) | 3.78 (1.68) | 5.26 (0.83) | 3.76 (1.56) |
| | Absent | Graphic | 11.35 (3.37) | 4.24 (1.32) | 1.23 (0.50) | | | | |
| | ($n$ = 46) | Text | 9.98 (2.68) | 4.20 (1.14) | 1.61 (0.72) | | | | |
| | | Overall | 10.66 (2.49) | 4.22 (1.04) | 1.42 (0.50) | 2.63 (1.38) | 4.04 (1.39) | 5.16 (0.72) | 4.41 (1.46) |
| | Overall | Graphic | 11.10 (3.42) | 4.07 (1.29) | 1.23 (0.49) | | | | |
| | ($n$ = 92) | Text | 10.10 (2.57) | 4.28 (1.20) | 1.51 (0.70) | | | | |
| | | Overall | 10.60 (2.53) | 4.18 (1.08) | 1.37 (0.48) | 2.52 (1.37) | 3.91 (1.54) | 5.21 (0.77) | 4.09 (1.54) |
| Overall | Present | Graphic | 11.30 (3.40) | 4.11 (1.19) | 1.22 (0.46) | | | | |
| | ($n$ = 92) | Text | 10.14 (2.71) | 4.41 (1.21) | 1.36 (0.59) | | | | |
| | | Overall | 10.72 (2.57) | 4.26 (1.03) | 1.29 (0.43) | 2.70 (1.56) | 4.10 (1.62) | 5.16 (0.87) | 3.98 (1.54) |
| | Absent | Graphic | 11.27 (3.40) | 4.25 (1.26) | 1.21 (0.50) | | | | |
| | ($n$ = 92) | Text | 10.21 (2.56) | 4.24 (1.23) | 1.45 (0.61) | | | | |
| | | Overall | 10.74 (2.45) | 4.25 (1.05) | 1.33 (0.45) | 2.66 (1.43) | 4.21 (1.34) | 5.16 (0.78) | 4.50 (1.35) |
| | Overall | Graphic | 11.29 (3.39) | 4.18 (1.23) | 1.21 (0.48) | | | | |
| | ($N$ = 184) | Text | 10.17 (2.63) | 4.33 (1.22) | 1.40 (0.60) | | | | |
| | | Overall | 10.73 (2.50) | 4.25 (1.04) | 1.31 (0.44) | 2.68 (1.50) | 4.16 (1.49) | 5.16 (0.82) | 4.24 (1.46) |

*Note.* The left part of the table (video-specific measures) refers to 2x2x2-factorial ANOVAs, the right part (cross-video measures) refers to the 2x2-factorial ANOVAs.

$F < 3.82$, $p > .051$.

### 3.3.2. Preregistered analyses

*3.3.2.1. Did the talking head negatively affect learning outcomes for certain materials?.* Contrary to Hypotheses 1 and 2.2, we found no significant three-way interaction Talking Head × Slide Type × Presentation Type for learning outcomes, $F(1, 180) = 2.73$, $p = .100$, $\eta_p^2 = 0.015$. However, the analysis revealed higher scores for the videos with graphic slides ($M = 11.29$, $SD = 3.39$) than for those with text slides ($M = 10.17$, $SD = 2.63$), $F(1, 180) = 19.36$, $p < .001$, $\eta_p^2 = 0.097$. Contrary to Hypothesis 2.1, there also was no interaction Slide Type × Presentation Type, $F < 1$. Further, there were no main effects of talking head or presentation type, and no other two-way interactions, all $F < 1$. In summary, our hypotheses on learning outcomes were not supported. Thus, regardless of slide type and presentation type, the talking head did not affect learning outcomes. For the conditions which corresponded to those of Experiment 1 (sequential presentation), we replicated the results of Experiment 1 finding no interplay of talking head and slide type.

Since we did not observe a main effect of presentation type or an interaction between slide type and presentation type, our results are not consistent with previous research (Fiorella et al., 2019; Jamet et al., 2008) reporting a beneficial effect of sequential presentation for graphic-based contents. Potentially, this inconsistency could be explained by the complexity of the learning materials because the graphic-based learning content may not have been visually complex enough for the sequential fade-in to provide an advantage over the static presentation (see also General Discussion).

*3.3.2.2. Did the talking head affect subjective ratings of the learning materials?.* For perceived learning ($M = 4.25$, $SD = 1.04$), in line with Hypothesis 3 and the results of Experiment 1, there was no main effect of the talking head, $F < 1$. Further, there were no other main effects or interactions, all $F < 2.48$, all $p > .117$. Also, in line with Hypothesis 3, the ANOVA on social presence ($M = 2.68$, $SD = 1.50$) showed no main effect of talking head, $F < 1$, and no other main effect or interaction, both $F < 2.11$, both $p > .148$. Potential explanations for the talking head not affecting ratings of perceived learning and social presence have already been addressed in Section 2.3.2.2.

### 3.3.3. Exploratory analyses

*3.3.3.1. Did the talking head affect intrinsic, germane, or extraneous cognitive load?.* Regarding intrinsic cognitive load, ratings were higher for videos with sequential presentation ($M = 4.40$, $SD = 1.40$) than for those with static presentation ($M = 3.91$, $SD = 1.54$), $F(1, 180) = 5.04$, $p = .026$, $\eta_p^2 = 0.027$. The main effect of talking head and the interaction Talking Head × Presentation Type were not significant, both $F < 1$.

For germane cognitive load ($M = 5.16$, $SD = 0.82$), the ANOVA revealed neither main effects of talking head and presentation type nor an interaction, all $F < 1$. Regarding extraneous cognitive load, participants rated extraneous load higher in conditions without talking head ($M = 4.50$, $SD = 1.35$) than in the conditions with talking head ($M = 3.98$, $SD = 1.54$), $F(1, 180) = 5.85$, $p = .017$, $\eta_p^2 = 0.031$. There was no main effect of presentation type and no interaction, both $F < 2.06$, both $p > .153$. The observed decrease in extraneous load due to a talking head, which we did not find in Experiment 1, appears counterintuitive, since features which are not directly related to the task should actually increase extraneous cognitive load (Sweller et al., 1998). However, our findings on extraneous load replicate the results of Wang, Antonenko, Keil, and Dawson (2020).

## 4. General discussion

### 4.1. Empirical contributions

By varying the presence of a talking head in addition to other features of educational videos with narrated slides (slide type, presentation type), this study is one of the first that systematically investigated potential boundary conditions of talking heads across different topics. In both experiments, our hypotheses, which assumed learning losses due to

the talking head with certain learning materials, were not supported. Hence, regardless of slide type and presentation type of the contents, we found no effects of the talking head on learning outcomes and thus identified no boundary conditions of instructor presence. Hence, the inclusion of a talking head offers neither clear advantages nor disadvantages. This lack of effects is consistent with some other studies investigating instructor presence in general (Homer et al., 2008, van Wermeskerken et al., 2018) and talking heads in particular (Kizilcec et al., 2014; Ng & Przybylek, 2021; Rosenthal & Walker, 2020). However, our study extends these findings by showing that a talking head has neither positive nor negative effects even if we systematically vary other design features such as (1) graphic-based or text-based slides and their (2) sequential or static presentation.

### 4.2. Theoretical implications

The lack of beneficial effects of the talking head on learning appears plausible in light of the embodiment principle (Mayer, 2021). In our experiments, the instructor was characterized by a low level of embodiment and did not use any gestures, whereas high embodiment (e.g., by providing gestures) has already proven to be beneficial for learning (Beege et al., 2020; Pi et al., 2017). Also consistent with the embodiment principle, the talking head did not affect social presence ratings in either experiment (see also Ng & Przybylek, 2021; Yuan et al., 2021). Additionally, our findings correspond to the image principle of multimedia learning (Mayer, 2021), which does not suggest any impact on learning with a rather static image of an instructor.

Since we did not observe detrimental effects of the talking head on learning outcomes, our results do not provide clear evidence for the coherence principle and the split-attention principle of multimedia learning (Mayer, 2021). It is possible that our materials were not complex enough to elicit learning differences. This rationale is also plausible considering that, in contrast to previous studies (Fiorella et al., 2019; Jamet et al., 2008), we did not observe a beneficial effect of the sequential presentation of graphic-based content on learning outcomes in Experiment 2. However, except for rather low difficulty ratings in both experiments (see Online Appendix), there is little support for this reasoning in our data (e.g., no ceiling effects in the knowledge test). Nevertheless, for future research investigating potential boundary conditions of talking heads, it would be interesting to examine learning videos which systematically vary in their visual complexity.

Additionally, it would also be conceivable that learning-impeding effects of a talking head postulated by coherence or split-attention principle (Mayer, 2021) were levelled out by beneficial effects of a talking head. However, neither does our data provide any direct indication for this, for example in terms of higher social presence ratings, nor does this reasoning appear particularly feasible due to the talking head's low level of embodiment in the present experiments.

Beyond, it is possible that research often does not identify main effects of talking heads on learning outcomes because talking heads are not equally effective (or distracting) for all different kinds of learners. In particular, learners' individual characteristics that were beyond the scope of this manuscript may play a role. Although this reasoning does not directly argue against the theories presented in this manuscript, it may be worthwhile to select a more systematic approach towards considering a broader range of individual prerequisites that may moderate how specific characteristics of the learning materials affect learning outcomes. For instance, initial work investigating the interplay of sustained attention and talking heads reports an interaction between both variables. Specifically, descriptive data suggest that learning-enhancing effects of the talking head are smaller for low than for high sustained attention, implying that sustained attention may influence talking head effects (Kokoç et al., 2020). Hence, since there often seems to be no one-size-fits-all solution, it would be desirable to incorporate individual prerequisites (e.g., attention span) more strongly into both research and theoretical models.

### 4.3. Limitations

The aforementioned findings should be interpreted in the light of some methodological choices, potential issues with differences in subjective prior knowledge, and scale reliabilities. With regard to methodological choices, we decided to use graphic-based and text-based videos covering different topics. This decision resulted from our intention to use ecologically valid materials which learners may actually encounter in real online learning scenarios. Thus, we opted against producing rather artificial videos which portray the same topic both graphic-based and text-based. Therefore, we do not interpret any main effects of slide type, as any resulting differences could have numerous reasons based on the different video contents. However, this does not affect the interpretation of potential main effects of the talking head as well as potential interactions of talking head and slide type because the comparisons of the different talking head conditions within the different levels of slide type are based on content equivalent materials.

Further, our knowledge test only measured factual knowledge and did not include questions addressing higher order learning (i.e., transfer knowledge). Therefore, all conclusions on learning outcomes refer to acquisition of factual knowledge and cannot be generalised to other learning measures. However, we specifically chose to focus on factual knowledge because we were mainly interested in potential learning-impeding effects of instructor presence (i.e., distraction). This should affect learners' extraction of factual knowledge from the learning materials, for example in situations in that they do not attend to the actual contents of the slides but focus on the instructor instead. Based on social agency theory (Mayer, 2014), it may have been interesting to include measures of transfer knowledge because this theory would suggest a positive effect of instructor presence on higher order learning outcomes. However, including measures for higher order learning was beyond the scope of the experiments due to their focus on potential boundary conditions of instructors (i.e., distracting effects) when extracting factual knowledge from narrated slides.

As a final methodological limitation, both experiments were conducted online, which comes with certain caveats. Indeed, online experiments offer less experimental control (over whether participants complete the study as intended) than laboratory experiments. We tried to address this limitation with various measures. For example, we included a control question to ensure that participants read the questions carefully and limited the time to answer the knowledge questions to prevent participants from looking up the correct answers. On a positive note, the online format also implies that our experiments are characterised by a high level of ecological validity because the learners watched the videos at home, which is closer to a real-world scenario (e.g., watching online lectures, YouTube videos) than watching them in a lab setting.

In addition, there were main effects of slide type on participants self-reported prior knowledge in both experiments. Even though we do not interpret the main effects of slide type as monocausal consequences of slide type on any dependent variable, the direction of the main effects regarding prior knowledge found in both experiments is opposite to the respective main effects regarding the knowledge test. Further, it should be noted, that subjective ratings of knowledge do not necessarily coincide with actual knowledge (Kardas & O'Brien, 2018; Toftness et al., 2018); which may be amplified by the fact that our assessment of prior knowledge was rather unspecific. However, we decided against using an objective prior knowledge test in order to avoid selectively directing learners' attention to the contents queried in the prior knowledge test.

Finally, it should be mentioned that the reliabilities of some variables (e.g., prior knowledge, perceived learning) were rather low. This could be due to the fact that the scales mentioned above partly consisted of only two items, whereas reliability coefficients tend to increase with many items (Field, 2009). In addition, the low reliability value could arise from the heterogeneity of the measured topics (Field, 2009; Stadler et al., 2021), because the two items of the scales asked for assessments of

two different videos with heterogeneous topics.

### 4.4. Practical implications and conclusion

In summary, including a talking head into educational videos with narrated slides had neither detrimental nor beneficial effects on knowledge acquisition. Our hypotheses regarding potential learning-impeding effects of the talking head were not supported in either experiment and we did not identify any boundary conditions for effects of instructor presence. Hence, when facing the question of whether to include a talking head in an educational video, our findings do not provide any clear evidence that this is particularly useful or beneficial for learning. This also implies that the inclusion of a visible talking head entails no disadvantages, even when presented next to materials that require visual attention. Our study supports the existing literature on talking heads and extends it by revealing no talking head effects even with systematically varied materials. However, considering theoretical assumptions and other empirical findings (i.e., high visual attention on the talking head), detrimental effects of the talking head on learning in certain constellations cannot be completely ruled out.

## CRediT authorship contribution statement

**Christina Sondermann:** Writing – original draft, Writing – review & editing, Conceptualization, Methodology, Project administration, Investigation, Formal analysis, Data curation. **Martin Merkt:** Writing – review & editing, Conceptualization, Methodology, Project administration, Supervision.

## 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

A link to data/code is included in the manuscript.

## References


Abrams, R. A., & Christ, S. E. (2003). Motion onset captures attention. *Psychological Science, 14*, 427–432. https://doi.org/10.1111/1467-9280.01458

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

Ayres, P., & Sweller, J. (2014). The split-attention principle in multimedia learning. In R. E. Mayer (Ed.), *The Cambridge handbook of multimedia learning* ((2nd ed.,, pp. 206–226). Cambridge University Press.

Baddeley, A. (1992). Working memory. *Science, 255*, 556–559. https://doi.org/10.1126/science.1736359

Beege, M., Ninaus, M., Schneider, S., Nebel, S., Schlemmel, J., Weidenmüller, J., … Rey, G. D. (2020). Investigating the effects of beat and deictic gestures of a lecturer in educational videos. *Computers & Education, 156*, Article 103955. https://doi.org/10.1016/j.compedu.2020.103955

Beege, M., Schneider, S., Nebel, S., & Rey, G. D. (2017). Look into my eyes!: Exploring the effect of addressing in educational videos. *Learning and Instruction, 49*, 113–120. https://doi.org/10.1016/j.learninstruc.2017.01.004

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*. Lawrence Erlbaum Associates.

Colliot, T., & Jamet, É. (2018). Understanding the effects of a teacher video on learning from a multimedia document: An eye-tracking study. *Educational Technology Research and Development, 66*, 1415–1433. https://doi.org/10.1007/s11423-018-9594-x

Eisinga, R., Grotenhuis, M. t., & Pelzer, B. (2013). The reliability of a two-item scale: Pearson, Cronbach, or Spearman-Brown? *International Journal of Public Health, 58*, 637–642. https://doi.org/10.1007/s00038-012-0416-3.

Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G*power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods, 39*, 175–191. https://doi.org/10.3758/BF03193146

Field, A. P. (2009). *Discovering statistics using SPSS* ((3rd ed.).). SAGE Publications.

Fiorella, L., Stull, A. T., Kuhlmann, S., & Mayer, R. E. (2019). Instructor presence in video lectures: The role of dynamic drawings, eye contact, and instructor visibility. *Journal of Educational Psychology, 111*, 1162–1171. https://doi.org/10.1037/edu0000325

Fiorella, L., & Mayer, R. E. (2016). Effects of observing the instructor draw diagrams on learning from multimedia messages. *Journal of Educational Psychology, 108*(4), 528–546. https://doi.org/10.1037/edu0000065

Guo, P. J., Kim, J., & Rubin, R. (2014). How video production affects student engagement. In M. Sahami, A. Fox, M. A. Hearst, & M. T.H. Chi (Eds.), *Proceedings of the first ACM conference on Learning @ scale conference - L@S '14* (pp. 41–50). ACM Press. https://doi.org/10.1145/2556325.2566239.

Hansch, A., Hillers, L., McConachie, K., Newman, C., Schildhauer, T., & Schmidt, P. (2015). Video and online learning: Critical reflections and findings from the field. *SSRN Electronic Journal.* Advance online publication. https://doi.org/10.2139/ssrn.2577882.

Hattie, J. (2008). Visible Learning. *Routledge*. https://doi.org/10.4324/9780203887332

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

Homer, B. D., Plass, J. L., & Blake, L. (2008). The effects of video on cognitive load and social presence in multimedia-learning. *Computers in Human Behavior, 24*, 786–797. https://doi.org/10.1016/j.chb.2007.02.009

Hong, J., Pi, Z., & Yang, J. (2018). Learning declarative and procedural knowledge via video lectures: Cognitive load and learning effectiveness. *Innovations in Education and Teaching International, 55*, 74–81. https://doi.org/10.1080/14703297.2016.1237371

Hoogerheide, V., Loyens, S. M., & van Gog, T. (2014). Comparing the effects of worked examples and modeling examples on learning. *Computers in Human Behavior, 41*, 80–91. https://doi.org/10.1016/j.chb.2014.09.013

Hoogerheide, V., van Wermeskerken, M., van Nassau, H., & van Gog, T. (2018). Model-observer similarity and task-appropriateness in learning from video modeling examples: Do model and student gender affect test performance, self-efficacy, and perceived competence? *Computers in Human Behavior, 89*, 457–464. https://doi.org/10.1016/j.chb.2017.11.012

Jamet, E., Gavota, M., & Quaireau, C. (2008). Attention guiding in multimedia learning. *Learning and Instruction, 18*, 135–145. https://doi.org/10.1016/j.learninstruc.2007.01.011

Kardas, M., & O'Brien, E. (2018). Easier seen than done: Merely watching others perform can foster an illusion of skill acquisition. *Psychological Science, 29*, 521–536. https://doi.org/10.1177/0956797617740646

Kizilcec, R. F., Bailenson, J. N., & Gomez, C. J. (2015). The instructor's face in video instruction: Evidence from two large-scale field studies. *Journal of Educational Psychology, 107*, 724–739. https://doi.org/10.1037/edu0000013

Kizilcec, R. F., Papadopoulos, K., & Sritanyaratana, L. (2014). Showing face in video instruction: Effects on information retention, visual attention, and affect. In M. Jones, P. Palanque, A. Schmidt, & T. Grossman (Eds.), *Proceedings of the 32nd annual ACM conference on Human factors in computing systems - CHI '14* (pp. 2095–2102). Association for Computing Machinery. https://doi.org/10.1145/2556288.2557207.

Klepsch, M., Schmitz, F., & Seufert, T. (2017). Development and validation of two instruments measuring intrinsic, extraneous, and germane cognitive load. *Frontiers in Psychology, 8*, 1997. https://doi.org/10.3389/fpsyg.2017.01997

Kokoç, M., Ilgaz, H., & Altun, A. (2020). Effects of sustained attention and video lecture types on learning performances. *Educational Technology Research and Development, 68*, 3015–3039. https://doi.org/10.1007/s11423-020-09829-7

Mayer, R. E. (2014). Principles based on social cues in multimedia learning: Personalization, voice, image and embodiment principles. In R. E. Mayer (Ed.), *The Cambridge handbook of multimedia learning* ((2nd ed.,, pp. 345–368). Cambridge University Press.

Mayer, R. E. (2021). *Multimedia learning* (3rd ed.). Cambridge University Press. https://doi.org/10.1017/9781316941355.

Mayer, R. E., & Fiorella, L. (2014). Principles for reducing extraneous processing in multimedia learning: Coherence, signaling, redundancy, spatial contiguity, and temporal contiguity principles. In R. E. Mayer (Ed.), *The Cambridge handbook of multimedia learning* (2nd ed., pp. 279–315). Cambridge University Press.

Mayer, R. E., & Massa, L. J. (2003). Three facets of visual and verbal learners: Cognitive ability, cognitive style, and learning preference. *Journal of Educational Psychology, 95*, 833–846. https://doi.org/10.1037/0022-0663.95.4.833

Moon, J., & Ryu, J. (2021). The effects of social and cognitive cues on learning comprehension, eye-gaze pattern, and cognitive load in video instruction. *Journal of Computing in Higher Education, 33*, 39–63. https://doi.org/10.1007/s12528-020-09255-x

Moreno, R., Mayer, R. E., Spires, H. A., & Lester, J. C. (2001). The Case for Social Agency in Computer-Based Teaching: Do Students Learn More Deeply When They Interact With Animated Pedagogical Agents? *Cognition and Instruction, 19*, 177–213. https://doi.org/10.1207/S1532690XCI1902_02

Ng, Y. Y., & Przybylek, A. (2021). Instructor Presence in Video Lectures: Preliminary Findings From an Online Experiment. *IEEE Access, 9*, 36485–36499. https://doi.org/10.1109/ACCESS.2021.3058735

Ouwehand, K., van Gog, T., & Paas, F. (2015). Designing effective video-based modeling examples using gaze and gesture cues. *Educational Technology & Society, 18*, 78–88.

Pi, Z., & Hong, J. (2016). Learning process and learning outcomes of video podcasts including the instructor and PPT slides: A Chinese case. *Innovations in Education and Teaching International, 53*, 135–144. https://doi.org/10.1080/14703297.2015.1060133

Pi, Z., Hong, J., & Yang, J. (2017). Effects of the instructor's pointing gestures on learning performance in video lectures. *British Journal of Educational Technology, 48*, 1020–1029. https://doi.org/10.1111/bjet.12471

Rosenthal, S., & Walker, Z. (2020). Experiencing Live Composite Video Lectures: Comparisons with Traditional Lectures and Common Video Lecture Methods. *International Journal for the Scholarship of Teaching and Learning, 14*, 1–10. https://doi.org/10.20429/ijsotl.2020.140108.

Schneider, S., Beege, M., Nebel, S., Schnaubert, L., & Rey, G. D. (2022). The Cognitive-Affective-Social Theory of Learning in digital Environments (CASTLE). *Educational Psychology Review, 34*, 1–38. https://doi.org/10.1007/s10648-021-09626-5

Schramm, H., & Hartmann, T. (2008). The PSI-Process Scales. A new measure to assess the intensity and breadth of parasocial processes. *Communications, 33*, 385–401. https://doi.org/10.1515/COMM.2008.025

Short, J., Williams, E., & Christie, B. (1976). *The social psychology of telecommunications*. Wiley.

Stadler, M., Sailer, M., & Fischer, F. (2021). Knowledge as a formative construct: A good alpha is not always better. *New Ideas in Psychology, 60*, Article 100832. https://doi.org/10.1016/j.newideapsych.2020.100832

Stull, A. T., Fiorella, L., & Mayer, R. E. (2018). An eye-tracking analysis of instructor presence in video lectures. *Computers in Human Behavior, 88*, 263–272. https://doi.org/10.1016/j.chb.2018.07.019

Stull, A. T., Fiorella, L., & Mayer, R. E. (2021). The case for embodied instruction: The instructor as a source of attentional and social cues in video lectures. *Journal of Educational Psychology., 113*, 1441–1453. https://doi.org/10.1037/edu0000650

Sweller, J. (1999). *Instructional design in technical areas*. ACER Press.

Sweller, J., Ayres, P., & Kalyuga, S. (2011). *Cognitive load theory*. Springer.

Sweller, J., van Merrienboer, J. J., & Paas, F. G. C. (1998). Cognitive architecture and instructional design. *Educational Psychology Review, 10*, 251–296. https://doi.org/10.1023/A:1022193728205

Toftness, A. R., Carpenter, S. K., Geller, J., Lauber, S., Johnson, M., & Armstrong, P. I. (2018). Instructor fluency leads to higher confidence in learning, but not better learning. *Metacognition and Learning, 13*, 1–14. https://doi.org/10.1007/s11409-017-9175-0

van Gog, T. (2014). The signaling (or cueing) principle in multimedia learning. In R. E. Mayer (Ed.), *The Cambridge handbook of multimedia learning* ((2nd ed.,, pp. 263–278). Cambridge University Press.

van Gog, T., & Rummel, N. (2010). Example-Based Learning: Integrating Cognitive and Social-Cognitive Research Perspectives. *Educational Psychology Review, 22*, 155–174. https://doi.org/10.1007/s10648-010-9134-7

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

van Wermeskerken, M., Ravensbergen, S., & van Gog, T. (2018). Effects of instructor presence in video modeling examples on attention and learning. *Computers in Human Behavior, 89*, 430–438. https://doi.org/10.1016/j.chb.2017.11.038

Wang, J., & Antonenko, P. (2017). Instructor presence in instructional video: Effects on visual attention, recall, and perceived learning. *Computers in Human Behavior, 71*, 79–89. https://doi.org/10.1016/j.chb.2017.01.049

Wang, J., Antonenko, P., & Dawson, K. (2020). Does visual attention to the instructor in online video affect learning and learner perceptions? An eye-tracking analysis. *Computers & Education, 146*, Article 103779. https://doi.org/10.1016/j.compedu.2019.103779

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*, 279–291. https://doi.org/10.1111/mbe.12239

Wilson, K. E., Martinez, M., Mills, C., D'Mello, S., Smilek, D., & Risko, E. F. (2018). Instructor presence effect: Liking does not always lead to learning. *Computers & Education, 122*, 205–220. https://doi.org/10.1016/j.compedu.2018.03.011

Yu, Z. (2021). The effect of teacher presence in videos on intrinsic cognitive loads and academic achievements. *Innovations in Education and Teaching International, 59*, 574–585. https://doi.org/10.1080/14703297.2021.1889394

Yuan, M., Zeng, J., Wang, A., & Shang, J. (2021). Would It Be Better if Instructors Technically Adjust Their Image or Voice in Online Courses? Impact of the Way of Instructor Presence on Online Learning. *Frontiers in Psychology, 12*, Article 746857. https://doi.org/10.3389/fpsyg.2021.746857

## Online Appendix (Manuscript 1)

The online appendix is part of the following manuscript:
Sondermann, C., & Merkt, M. (2023). What is the effect of talking heads in educational videos with different types of narrated slides? *Contemporary Educational Psychology*, *74*, Article 102207. https://doi.org/10.1016/j.cedpsych.2023.102207

*Note: The online appendix has been adapted to the formalities and numbering of the present thesis, without changing the content. The original online appendix can be downloaded from OSF at the following link or by scanning the QR code below:*
https://osf.io/xug9c/?view_only=7b149959778c40c3b9fe51fb988f823f

**Content**

Additional measures and results

## Additional Measures and Results

### Additional Measures Experiment 1a

#### *Prior Interest*

We asked for participants' prior interest in the six topics (e.g., "How much are you interested in the African country Burkina Faso?") on a 7-point Likert scale from 1 (*not at all*) to 7 (*very interested*). Each participant's ratings were averaged to get an a priori interest score for graphic- (α = .62) and text-based topics (α = .57).

#### *Learning Preference*

Since some previous studies showed an interplay between instructor presence and learning preference (Homer et al., 2008; Kizilcec et al., 2015), we assessed learning preference as a control variable the ensure the comparability of the between-subjects conditions. We used the one-item Verbal-Visual Learning Style Rating (Mayer & Massa, 2003). Learners rated their preference for visual or verbal content on a 7-point scale.

#### *Satisfaction, Professionalism, and Joy of Learning*

Learners rated their satisfaction[4] with the learning materials ("How satisfied are you with the learning videos?", 1 = *very dissatisfied*, 7 = *very satisfied*, adapted from Wang and Antonenko, 2017), the professionalism of the videos ("How professional did the videos appear?", 1 = *very unprofessional*, 7 = *very professional*, adapted from Merkt et al., 2020), and their joy of learning ("How much did you enjoy learning with the videos?", 1 = *not at all*, 7 = *a lot*, adapted from Merkt et al., 2020) on 7-point Likert scales.

#### *Invested Mental Effort*

Learners rated their mental effort ("How exhausting did you find it to follow the video content?", adapted from Paas, 1992) on a 9-point Likert scale ranging from 1 (*not at all exhausting*) to 9 (*extremely exhausting*). Mean values were calculated for graphic-based (α = .53) and text-based videos (α = .76).

#### *Perceived Difficulty*

Learners rated perceived difficulty ("How easy or difficult was this video to understand?", Kalyuga et al., 1999) on a 7-point scale from 1 (*extremely easy*) to 7 (*extremely difficult*). We calculated mean values for graphic-based (α = .59) and text-based videos (α = .68).

### Additional Results Experiment 1a

An overview of the results of Experiment 1a can be found in Table B.1. Descriptive data on all variables can be found in Table B.2.

[4] Based on prior research, we expected higher satisfaction ratings with than without talking head.

***Mental Effort and Perceived Difficulty***

For mental effort, results indicated no main effects of slide type, $F(1, 94) = 3.26$, $p = .074$, $\eta_p^2 = .033$, or talking head, $F < 1$, and no interaction of the two variables, $F < 1$. For perceived difficulty, we found higher difficulty ratings for graphic-based videos ($M = 3.29$, $SD = 1.28$) than for text-based videos ($M = 2.70$, $SD = 1.07$), $F(1, 94) = 30.11$, $p < .001$, $\eta_p^2 = .243$. However, the main effect of talking head and the interaction of both variables were not significant, both $F < 1$.

***Satisfaction, Professionalism, and Joy of Learning***

For satisfaction, the ANOVA revealed no main effect of talking head, $F(1, 94) = 1.53$, $p = .220$, $\eta_p^2 = .016$. Further, there was no significant effect of talking head on either professionalism, $F(1, 94) = 1.83$, $p = .179$, $\eta_p^2 = .019$ or joy of learning, $F < 1$.

***Prior Interest and Learning Preference***

For interest, there was no main effect of either slide type, $F(1, 94) = 1.35$, $p = .249$, $\eta_p^2 = .014$, or talking head, $F < 1$, and no interaction, $F < 1$. With regard to learning preference, there was no main effect of talking head condition, $F(1, 94) = 1.95$, $p = .165$, $\eta_p^2 = .020$, indicating that between-subjects conditions did not differ in terms of prior interest and learning preference.

**Additional Measures Experiment 1b**

The variables satisfaction, professionalism, joy of learning[5], invested mental effort ($\rho = .41$ for graphic-based, $\rho = .75$ for text-based videos), and perceived difficulty ($\rho = .46$ for graphic-based, $\rho = .65$ for text-based videos), and cognitive style were collected identical to Experiment 1a. Also, prior interest for the four topics was recorded as in Experiment 1a. Reliabilities were $\rho = .41$ for interest in graphic-based topics, and $\rho = .51$ for interest in text-based topics.

**Additional Results Experiment 1b**

An overview of the results of Experiment 1b can be found in Table B.3. Descriptive data on all variables can be found in Table B.4.

***Mental Effort and Perceived Difficulty***

For mental effort, there were no main effects of talking head, slide type, and presentation type, all $F < 1$. Further, there were no significant interactions Talking Head x Slide Type x Presentation Type, $F(1, 180) = 2.76$, $p = .098$, $\eta_p^2 = .015$, Talking Head x Slide Type, $F < 1$,

[5] Based on the results of Experiment 1a, we expected no differences between conditions with and without talking head for ratings of satisfaction, professionalism, and joy of learning.

**Table B.1**

| | Mental effort | Perceived difficulty | Prior interest | Satisfaction | Professionalism | Joy of learning | Learning preference |
|---|---|---|---|---|---|---|---|
| Main effect talking head | $F < 1$ | $F < 1$ | $F < 1$ | $F(1, 94) = 1.53$, $p = .220$, $\eta_p^2 = .016$ | $F(1, 94) = 1.83$, $p = .179$, $\eta_p^2 = .019$ | $F < 1$ | $F(1, 94) = 1.95$, $p = .165$, $\eta_p^2 = .020$ |
| Main effect slide type [a] | $F(1, 94) = 3.26$, $p = .074$, $\eta_p^2 = .033$ | $F(1, 94) = 30.11$, $p < .001$, $\eta_p^2 = .243$ | $F(1, 94) = 1.35$, $p = .249$, $\eta_p^2 = .014$ | | | | |
| Talking Head x Slide Type | $F < 1$ | $F < 1$ | $F < 1$ | | | | |

*An overview of the ANOVA results from Experiment 1a*

[a] For completeness, main effects of slide type are reported but not interpreted due to different video contents.

**Table B.2**

*Means (and Standard Deviations in Parentheses) for Experiment 1a*

| Condition | | Video-specific measures | | | Cross-video measures | | | |
|---|---|---|---|---|---|---|---|---|
| Instructor | Slide type | Mental effort | Perceived difficulty | Prior interest | Satisfaction | Professionalism | Joy of learning | Learning preference |
| Present ($n = 48$) | Graphic | 5.22 (1.75) | 3.23 (1.19) | 2.77 (1.28) | | | | |
| | Text | 4.74 (1.92) | 2.64 (1.08) | 2.84 (1.15) | | | | |
| | Overall | 4.98 (1.56) | 2.93 (1.01) | 2.81 (1.11) | 3.60 (1.54) | 3.73 (1.48) | 3.25 (1.58) | 4.23 (1.13) |
| Absent ($n = 48$) | Graphic | 5.08 (1.86) | 3.35 (1.38) | 2.63 (1.25) | | | | |
| | Text | 4.86 (1.92) | 2.77 (1.07) | 2.79 (1.06) | | | | |
| | Overall | 4.97 (1.65) | 3.06 (1.12) | 2.71 (1.07) | 3.23 (1.43) | 3.31 (1.53) | 3.21 (1.52) | 4.56 (1.20) |
| Overall ($N = 96$) | Graphic | 5.15 (1.80) | 3.29 (1.28) | 2.70 (1.26) | | | | |
| | Text | 4.80 (1.91) | 2.70 (1.07) | 2.82 (1.10) | | | | |
| | Overall | 4.98 (1.60) | 3.00 (1.06) | 2.76 (1.08) | 3.42 (1.49) | 3.52 (1.51) | 3.23 (1.54) | 4.40 (1.17) |

*Note.* The left part of the table (video-specific measures) refers to 2x2-factorial ANOVAs, the right part (cross-video measures) refers to the one-way ANOVAs.

Presentation Type, $F(1, 180) = 2.76$, $p = .098$, $\eta_p^2 = .015$, Talking Head x Slide Type, $F < 1$, and Slide Type x Presentation Type, $F(1, 180) = 3.46$, $p = .065$, $\eta_p^2 = .019$, and Talking Head x Presentation Type, $F(1, 180) = 1.03$, $p = .311$, $\eta_p^2 = .006$.

For perceived difficulty, the ANOVA revealed a three-way interaction Talking Head x Slide Type x Presentation Type, $F(1, 180) = 4.64$, $p = .033$, $\eta_p^2 = .025$, and a significant two-way interaction Slide Type x Presentation Type, $F(1, 180) = 3.95$, $p = .048$, $\eta_p^2 = .021$. The interactions Talking Head x Slide Type, $F(1, 180) = 1.73$, $p = .190$, $\eta_p^2 = .010$, and Talking Head x Presentation Type, $F < 1$, were not significant. Further, there were higher difficulty ratings for graphic-based videos ($M = 3.45$, $SD = 1.34$) than for text-based videos ($M = 2.59$, $SD = 1.09$), $F(1, 180) = 78.69$, $p < .001$, $\eta_p^2 = .304$. Effects of talking head and presentation type were not significant, both $F < 1$.

To resolve the three-way interaction, we followed two different approaches reflecting (1) the replication and extension of the findings of Experiment 1a and (2) the question whether the talking head qualifies the two-way interaction Slide Type x Presentation Type observed in the overall ANOVA.

First, to replicate and extend the findings of Experiment 1a, we conducted separate 2x2-factorial ANOVAs with the factors talking head and slide type for the two different presentation types. For sequential presentation, (the same presentation type as in Experiment 1a), the findings could be replicated. In particular, there were higher values of perceived difficulty for graphic slides ($M = 3.33$, $SD = 1.40$) than for text slides ($M = 2.66$, $SD = 1.15$), $F(1, 90) = 22.56$, $p < .001$, $\eta_p^2 = .200$, whereas the interaction Talking Head x Slide Type, $F < 1$, as well as the main effect of talking head, $F(1, 90) = 1.27$, $p = .264$, $\eta_p^2 = .014$, were not significant. For the static presentation condition, a different pattern of results emerged: Again, there were higher values of perceived difficulty for graphic slides ($M = 3.57$, $SD = 1.26$) compared to text slides ($M = 2.52$, $SD = 1.04$), $F(1, 90) = 62.07$, $p < .001$, $\eta_p^2 = .408$, and no main effect of talking head, $F < 1$. Importantly, the analysis also showed a significant interaction Talking Head x Slide Type, $F(1, 90) = 6.34$, $p = .014$, $\eta_p^2 = .066$. When no talking head was present, Bonferroni-adjusted post-hoc comparisons revealed that videos with graphic slides ($M = 3.40$, $SD = 1.20$) were perceived as more difficult than videos with text slides ($M = 2.68$, $SD = 1.21$), $F(1, 90) = 14.37$, $p < .001$, $\eta_p^2 = .138$. When the talking head was present, participants also perceived the graphic slides ($M = 3.74$, $SD = 1.31$) to be more difficult than the text slides ($M = 2.35$, $SD = 0.81$), $F(1, 90) = 54.05$, $p < .001$, $\eta_p^2 = .375$. However, this effect was more than twice as large for videos with talking head.

**Table B.3**

*An overview of the ANOVA results from Experiment 1b*

| | Mental effort | Perceived difficulty | Prior interest | Satisfaction | Professionalism | Joy of learning | Learning preference |
|---|---|---|---|---|---|---|---|
| Main effect talking head | $F < 1$ | $F < 1$ | $F < 1$ | $F(1,180) = 2.28$, $p = .133$, $\eta_p^2 = .013$ | $F(1,180) = 4.09$, $p = .045$, $\eta_p^2 = .022$ | $F < 1$ | $F < 1$ |
| Main effect slide type [a] | $F < 1$ | $F(1,180) = 78.69$, $p < .001$, $\eta_p^2 = .304$ | $F(1, 180) = 15.49$, $p < .001$, $\eta_p^2 = .079$ | | | | |
| Main effect presentation type | $F < 1$ | $F < 1$ | $F < 1$ | $F < 1$ | $F(1,180) = 2.37$, $p = .125$, $\eta_p^2 = .013$ | $F < 1$ | $F < 1$ |
| Talking Head x Slide Type | $F < 1$ | $F(1, 180) = 1.73$, $p = .190$, $\eta_p^2 = .010$ | $F < 1$ | | | | |
| Slide Type x Presentation Type | $F(1,180) = 3.46$, $p = .065$, $\eta_p^2 = .019$ | $F(1, 180) = 3.95$, $p = .048$, $\eta_p^2 = .021$ | $F(1, 180) = 1.06$, $p = .305$, $\eta_p^2 = .006$ | | | | |
| Talking Head x Presentation Type, | $F(1,180) = 1.03$, $p = .311$, $\eta_p^2 = .006$ | $F < 1$ | $F(1, 180) = 1.21$, $p = .273$, $\eta_p^2 = .007$ | $F < 1$ | $F < 1$ | $F < 1$ | $F < 1$ |
| Talking Head x Slide Type x Presentation Type | $F(1,180) = 2.76$, $p = .098$, $\eta_p^2 = .015$ | $F(1, 180) = 4.64$, $p = .033$, $\eta_p^2 = .025$ | $F < 1$ | | | | |

[a] For completeness, main effects of slide type are reported but not interpreted due to different video contents.

**Table B.4**

*Means (and Standard Deviations in Parentheses) for Experiment* 1b

| Condition | | | Video-specific measures | | | Cross-video measures | | | |
|---|---|---|---|---|---|---|---|---|---|
| Presentation type | Instructor | Slide type | Mental effort | Perceived difficulty | Prior interest | Satisfaction | Professionalism | Joy of learning | Learning preference |
| Sequential | Present | Graphic | 4.65 (1.98) | 3.16 (1.36) | 2.76 (1.18) | | | | |
| | (*n* = 46) | Text | 5.10 (1.86) | 2.58 (1.08) | 3.33 (1.29) | | | | |
| | | Overall | 4.88 (1.48) | 2.87 (1.03) | 3.04 (1.08) | 3.80 (1.29) | 4.22 (1.58) | 3.41 (1.29) | 4.35 (1.10) |
| | Absent | Graphic | 5.24 (2.06) | 3.50 (1.44) | 2.64 (1.77) | | | | |
| | (*n* = 46) | Text | 5.38 (2.22) | 2.75 (1.21) | 2.97 (1.38) | | | | |
| | | Overall | 5.31 (1.72) | 3.13 (1.15) | 2.80 (1.09) | 3.39 (1.37) | 3.70 (1.56) | 3.50 (1.43) | 4.24 (1.20) |
| | Overall | Graphic | 4.95 (2.03) | 3.33 (1.40) | 2.70 (1.17) | | | | |
| | (*n* = 92) | Text | 5.24 (2.05) | 2.66 (1.15) | 3.15 (1.34) | | | | |
| | | Overall | 5.09 (1.61) | 3.00 (1.09) | 2.92 (1.08) | 3.60 (1.34) | 3.96 (1.58) | 3.46 (1.35) | 4.29 (1.10) |
| Static | Present | Graphic | 5.33 (1.92) | 3.74 (1.31) | 2.76 (1.29) | | | | |
| | (*n* = 46) | Text | 4.61 (2.08) | 2.35 (0.81) | 3.08 (1.32) | | | | |
| | | Overall | 4.97 (1.78) | 3.04 (0.88) | 2.92 (1.13) | 3.89 (1.34) | 3.80 (1.54) | 3.41 (1.53) | 4.33 (1.21) |
| | Absent | Graphic | 4.87 (1.81) | 3.40 (1.20) | 2.92 (1.44) | | | | |
| | (*n* = 46) | Text | 4.95 (1.99) | 2.68 (1.21) | 3.13 (1.05) | | | | |
| | | Overall | 4.91 (1.61) | 3.04 (1.02) | 3.03 (0.98) | 3.70 (1.46) | 3.41 (1.44) | 3.37 (1.55) | 4.37 (1.08) |
| | Overall | Graphic | 5.10 (1.87) | 3.57 (1.26) | 2.84 (1.22) | | | | |
| | (*n* = 92) | Text | 4.78 (2.03) | 2.52 (1.04) | 3.10 (1.19) | | | | |
| | | Overall | 4.94 (1.69) | 3.04 (0.95) | 2.97 (1.06) | 3.79 (1.39) | 3.61 (1.50) | 3.39 (1.53) | 4.35 (1.14) |
| Overall | Present | Graphic | 4.99 (1.97) | 3.45 (1.36) | 2.76 (1.23) | | | | |
| | (*n* = 92) | Text | 4.85 (1.98) | 2.46 (0.96) | 3.20 (1.30) | | | | |
| | | Overall | 4.92 (1.63) | 2.96 (0.96) | 2.98 (1.10) | 3.85 (1.31) | 4.01 (1.57) | 3.41 (1.41) | 4.34 (1.15) |
| | Absent | Graphic | 5.05 (1.94) | 3.45 (1.32) | 2.78 (1.16) | | | | |
| | (*n* = 92) | Text | 5.16 (2.11) | 2.72 (1.21) | 3.05 (1.22) | | | | |
| | | Overall | 5.11 (1.67) | 3.08 (1.08) | 2.92 (1.04) | 3.54 (1.42) | 3.55 (1.50) | 3.43 (1.48) | 4.30 (1.10) |
| | Overall | Graphic | 5.02 (1.95) | 3.45 (1.34) | 2.77 (1.19) | | | | |
| | (*n* = 184) | Text | 5.01 (2.05) | 2.59 (1.09) | 3.13 (1.26) | | | | |
| | | Overall | 5.01 (1.65) | 3.02 (1.02) | 2.95 (1.07) | 3.70 (1.37) | 3.78 (1.55) | 3.42 (1.44) | 4.32 (1.12) |

*Note.* The left part of the table (video-specific measures) refers to 2x2x2-factorial ANOVAs, the right part (cross-video measures) refers to the 2x2-factorial ANOVAs.

Second, to address the question whether the talking head qualifies the two-way interaction of slide type and presentation type observed in the overall ANOVA, we conducted two separate 2x2-factorial ANOVAs for the two talking head conditions. For conditions without talking head, there were higher difficulty ratings for graphic slides ($M = 3.45$, $SD = 1.32$) than for text slides ($M = 2.72$, $SD = 1.21$), $F(1, 90) = 28.44$, $p < .001$, $\eta_p^2 = .240$. There was no main effect of presentation type, $F < 1$, and no interaction Slide Type x Presentation Type, $F < 1$.

For conditions with talking head, the post-hoc ANOVA revealed higher difficulty ratings for graphic-based videos ($M = 3.45$, $SD = 1.36$) than for text-based videos ($M = 2.46$, $SD = 0.96$), $F(1, 90) = 52.06$, $p < .001$, $\eta_p^2 = .366$. There was no main effect of presentation type, $F < 1$, but a significant interaction Slide Type x Presentation Type, $F(1, 90) = 8.61$, $p = .004$, $\eta_p^2 = .087$. When resolving this two-way interaction, Bonferroni-adjusted post-hoc comparisons revealed no main effect of presentation type for videos with text slides, $F(1, 90) = 1.32$, $p = .254$, $\eta_p^2 = .014$. However, for videos with graphic slides, comparisons revealed that videos with static presentation ($M = 3.74$, $SD = 1.31$) were perceived as more difficult than videos with sequential presentation ($M = 3.16$, $SD = 1.36$), $F(1, 90) = 4.27$, $p = .042$, $\eta_p^2 = .045$. Thus, the combination of talking head, graphic slides, and static presentation produced the highest difficulty ratings. Additionally, the interaction Slide Type x Presentation Type can be resolved in a different way using Bonferroni-adjusted post-hoc comparisons: For sequential presentation, participants considered graphic-based videos ($M = 3.16$, $SD = 1.36$) as more difficult than text-based videos ($M = 2.58$, $SD = 1.08$), $F(1, 90) = 9.17$, $p = .003$, $\eta_p^2 = .092$. For static presentation, perceived difficulty was also significantly higher for graphic slides ($M = 3.74$, $SD = 1.31$) than for text slides ($M = 2.35$, $SD = 0.81$), $F(1, 90) = 51.50$, $p < .001$, $\eta_p^2 = .364$. However, the latter effect was about four times larger. Thus, sequential presentation produced smaller differences in difficulty perceptions between graphic- and text-based videos.

***Satisfaction, Professionalism, and Joy of Learning***

For satisfaction, in line with our hypothesis (OSF, link to preregistration excluded for blind reviews),the ANOVA revealed no main effect of talking head, $F(1, 180) = 2.28$, $p = .133$, $\eta_p^2 = .013$, and further no main effect of presentation type, $F < 1$, or an interaction Talking Head x Presentation Type, $F < 1$.

Contrary to Experiment 1a, participants perceived videos with talking head ($M = 4.01$, $SD = 1.57$) as more professional than videos without talking head ($M = 3.55$, $SD = 1.50$), $F(1, 180) = 4.09$, $p = .045$, $\eta_p^2 = .022$. There was no main effect of presentation type, $F(1, 180) = 2.37$, $p = .125$, $\eta_p^2 = .013$, and no significant interaction Talking Head x Presentation Type, $F < 1$.

Regarding joy of learning, results revealed neither a main effect of talking head (supporting our hypothesis) and presentation type, nor an interaction, all $F < 1$.

***Prior Interest and Learning Preference***

For interest, the ANOVA revealed higher values for the text-based topics ($M = 3.13$, $SD = 1.26$) than for graphic-based topics ($M = 2.77$, $SD = 1.19$), $F(1, 180) = 15.49$, $p < .001$, $\eta_p^2 = .079$. There were no main effects of talking head and presentation type, both $F < 1$. Further, there were no interactions Talking Head x Slide Type, $F < 1$, Slide Type x Presentation Type, $F(1, 180) = 1.06$, $p = .305$, $\eta_p^2 = .006$, Talking Head x Presentation Type, $F(1, 180) = 1.21$, $p = .273$, $\eta_p^2 = .007$, and Talking Head x Slide Type x Presentation Type, $F < 1$. The ANOVA for preference revealed no significant main effects of talking head or presentation type, and no significant interaction, all $F < 1$, suggesting no difference between the between-subjects conditions.

**References (Online Appendix)**

Homer, B. D., Plass, J. L., & Blake, L. (2008). The effects of video on cognitive load and social presence in multimedia-learning. *Computers in Human Behavior*, *24*, 786–797. https://doi.org/10.1016/j.chb.2007.02.009

Kalyuga, S., Chandler, P., & Sweller, J. (1999). Managing split-attention and redundancy in multimedia instruction. *Applied Cognitive Psychology*, *13*, 351–371. https://doi.org/10.1002/(SICI)1099-0720(199908)13:4<351::AID-ACP589>3.0.CO;2-6

Kizilcec, R. F., Bailenson, J. N., & Gomez, C. J. (2015). The instructor's face in video instruction: Evidence from two large-scale field studies. *Journal of Educational Psychology*, *107*, 724–739. https://doi.org/10.1037/edu0000013

Mayer, R. E., & Massa, L. J. (2003). Three facets of visual and verbal learners: Cognitive ability, cognitive style, and learning Preference. *Journal of Educational Psychology*, *95*, 833–846. https://doi.org/10.1037/0022-0663.95.4.833

Merkt, M., Lux, S., Hoogerheide, V., van Gog, T., & Schwan, S. (2020). A change of scenery: Does the setting of an instructional video affect learning? *Journal of Educational Psychology*, *112*, 1273–1283. https://doi.org/10.1037/edu0000414

Paas, F. G. (1992). Training strategies for attaining transfer of problem-solving skill in statistics: A cognitive-load approach. *Journal of Educational Psychology*, *84*, 429–434. https://doi.org/10.1037/0022-0663.84.4.429

Wang, J., & Antonenko, P. (2017). Instructor presence in instructional video: Effects on visual attention, recall, and perceived learning. *Computers in Human Behavior*, *71*, 79–89. https://doi.org/10.1016/j.chb.2017.01.049

# Appendix B

Manuscript 2

Sondermann, C. & Merkt, M. (2023). Like it or learn from it: Effects of talking heads in educational videos. *Computers & Education, 193*, Article 104675. https://doi.org/10.1016/j.compedu.2022.104675

| **Author** | **Author position** | **Scientific ideas %** | **Data generation %** | **Analysis and interpretation %** | **Paper writing %** |
|---|---|---|---|---|---|
| Christina Sondermann | 1 | 80 | 100 | 90 | 90 |
| Martin Merkt | 2 | 20 | 0 | 10 | 10 |

Computers & Education 193 (2023) 104675

Contents lists available at ScienceDirect

Computers & Education

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

# Like it or learn from it: Effects of talking heads in educational videos☆

Christina Sondermann *, Martin Merkt

*German Institute for Adult Education – Leibniz Centre for Lifelong Learning, Heinemannstraße 12-14, 53175, Bonn, Germany*

A R T I C L E I N F O

*Keywords:*
Distance education and online learning
Informal learning
Media in education

A B S T R A C T

When considering the inclusion of a visible instructor into educational videos, theory and prior research point to both advantages and disadvantages and suggest potential interactions with other design features of educational videos. Therefore, we conducted an online experiment in which $N = 112$ participants watched eight short videos with narrated slides on different topics, using a within-design to vary talking head (present vs. absent), slide type (graphic vs. text), and presentation type (sequential fade-in vs. static). We investigated effects of these design features on learning outcomes (factual knowledge), learners' ratings of the videos (e.g., perceived learning, satisfaction), and selection behaviour. Results revealed worse learning outcomes for videos with talking head, while participants rated their perceived learning higher for those videos with talking head. Further, participants indicated higher satisfaction for videos with talking head and selected them more frequently when choosing between different formats. Potential explanations for the findings and directions for future research are discussed.

## 1. Introduction

Initiated by digitalisation and most recently promoted by the COVID-19 pandemic, educational videos play a central role in formal and informal learning. However, there are still unanswered questions about how educational videos should be designed so that learners select them, like them, and successfully learn with them. One design feature of educational videos whose effects are still not well understood is the inclusion of a visible instructor. Instead of only hearing the instructor's voice, many videos include a video of the talking instructor next to the actual learning content. In our experiment, we investigate the mere presence of a talking head without further didactic functions (e.g., pointing, writing, gesturing) next to narrated slides. For research on instructors with further didactic functions, see for example Beege et al. (2020), Fiorella, Stull, Kuhlmann, and Mayer (2019), or Stull, Fiorella, and Mayer (2021). Theory argues for both potential advantages (e.g., social agency theory, Mayer, 2014a) and disadvantages (e.g., split-attention principle, Ayres & Sweller, 2014) of instructor presence. Similarly, research does not provide consistent findings regarding the inclusion of an instructor in educational videos.

☆ The study received an ethics approval from the local ethics committee (ethikkommission@die-bonn.de) at the German Institute for Adult Education - Leibniz Centre for Lifelong Learning (approval number: DIE-LEK 2020–007). Informed consent was obtained from all individual participants.

* Corresponding author.
*E-mail addresses:* sondermann@die-bonn.de (C. Sondermann), merkt@die-bonn.de (M. Merkt).

https://doi.org/10.1016/j.compedu.2022.104675
Received 26 July 2022; Received in revised form 2 November 2022; Accepted 4 November 2022
Available online 5 November 2022

### 1.1. Theoretical background

On the one hand, including the visible instructor in educational videos provides a social cue that can activate social interaction schemas. According to social agency theory (SAT, Mayer, 2014a), social cues can elicit social responses (e.g., attention), which in turn lead to deeper processing and enhanced transfer learning. Correspondingly, Schneider, Beege, Nebel, Schnaubert, and Rey (2022) highlight the relevance of social processes in learning with digital materials in their cognitive-affective-social theory of learning in digital environments (CASTLE), according to which a talking head can be considered a strong social cue. Further, following the social-cue hypothesis (Colliot & Jamet, 2018), social cues can translate not only into better learning outcomes, but also into an enhanced learning experience for learners (e.g., higher interest and engagement).

On the other hand, cognitive load theory (CLT, Sweller, 1999; Sweller, Ayres, & Kalyuga, 2011) and cognitive theory of multimedia learning (CTML, Mayer, 2014b) argue that extraneous materials should be excluded from the learning materials to prevent cognitive overload (see coherence principle, Mayer & Fiorella, 2014). Adding an instructor can be considered as extraneous material when the instructor does not directly contribute to the task itself (e.g., by performing further didactic functions such as pointing).

In addition, people show a preference for human faces (Gullberg & Holmqvist, 2006; Johnson, Dziurawiec, Ellis, & Morton, 1991). In line with this, several studies on instructor presence revealed that learners spent a significant amount of time looking at the instructor (e.g., Kizilcec, Papadopoulos, & Sritanyaratana, 2014; van Wermeskerken, Ravensbergen, & van Gog, 2018). Therefore, the instructor visually competes with the learning content. According to the split-attention principle (CTML, Ayres & Sweller, 2014), learning materials that require learners to split their attention can increase extraneous cognitive load and therefore hinder learning.

Taken together, the inclusion of a talking head may serve as social cue that enhances learning on the one hand (SAT, Mayer, 2014a), whereas the talking head can be seen as source of extraneous cognitive load that hinders learning on the other (CTML, Ayres & Sweller, 2014; Mayer & Fiorella, 2014).

### 1.2. Prior research on instructor presence

#### 1.2.1. Effects on learning outcomes

In line with SAT (Mayer, 2014a), several studies demonstrated enhanced learning through a visible instructor (e.g., Pi & Hong, 2016). However, these beneficial effects were often present in specific constellations of learning materials and/or types of knowledge, for instance, only for transfer performance (Wang, Antonenko, Keil, & Dawson, 2020), for recall performance in easy videos (Wang & Antonenko, 2017), for transfer performance in difficult videos (Wang, Antonenko, & Dawson, 2020), for declarative knowledge (Hong, Pi, & Yang, 2018), or for retention of spoken explanations (Colliot & Jamet, 2018). In contrast, several studies reported no learning differences between learners watching videos with and without an instructor, neither for immediate nor for delayed recall (Kizilcec et al., 2014; Ng & Przybylek, 2021), and neither for retention nor for transfer knowledge (Homer, Plass, & Blake, 2008; Sondermann & Merkt, 2022; van Wermeskerken et al., 2018).

Although some multimedia principles suggest inhibitory effects for learning and although the instructor attracted considerable visual attention in previous studies (e.g., Kizilcec et al., 2014; van Wermeskerken et al., 2018), to the best of our knowledge, no study unanimously revealed detrimental effects on learning outcomes when comparing videos with and without human instructor (see also Henderson & Schroeder, 2021; but see Wilson et al., 2018). However, research on the effects of instructor presence and its interplay with other design features such as slide type and presentation type suggested that learners perceived the combination of present instructor, graphic slides, and static presentation as particularly difficult (Sondermann & Merkt, 2022).

#### 1.2.2. Effects on subjective ratings

Considering the practical relevance, not only objective learning outcomes but also learners' subjective ratings of the learning materials are of central importance. As soon as individuals engage in non-supervised, free-choice informal learning, it becomes particularly important whether they like a video because their experience may also determine if they decide to watch or continue watching educational videos (Guo, Kim, & Rubin, 2014).

In line with the social-cue hypothesis (Colliot & Jamet, 2018), several studies revealed that including a visible instructor enhanced learners' perceived learning (Kizilcec, Bailenson, & Gomez, 2015; Wang & Antonenko, 2017; Wilson et al., 2018), satisfaction (Wang & Antonenko, 2017; Wang, Antonenko, & Dawson, 2020), and situational interest (Wang, Antonenko, & Dawson, 2020; Wilson et al., 2018). In contrast, other studies showed that a visible instructor did not affect social presence (Colliot & Jamet, 2018; Homer et al., 2008; Ng & Przybylek, 2021; Sondermann & Merkt, 2022), situational interest (Colliot & Jamet, 2018), satisfaction, or perceived learning (Sondermann & Merkt, 2022).

These contradictory findings may be partly attributed to the fact that those learners who reported a positive impact on video ratings (e.g., Kizilcec et al., 2014; Wang & Antonenko, 2017; Wilson et al., 2018) had a direct comparison between videos with and without instructors (e.g., because they watched both videos with and without instructor). In contrast, studies that found no effects on ratings (Sondermann & Merkt, 2022; Colliot & Jamet, 2018; Homer et al., 2008; Ng & Przybylek, 2021) varied instructor presence as a between-subjects factor (i.e., learners either watched only videos with or without instructor). Nevertheless, this methodological difference is only one potential explanation, because there are numerous differences between the various studies that may contribute to the heterogeneous findings (e.g., learning materials).

Similar to learning outcomes and affective ratings, research regarding the effects of instructor presence on cognitive load (CL) provided heterogeneous results: Although some multimedia principles (Ayres & Sweller, 2014; Mayer & Fiorella, 2014) suggest that a visible instructor results in higher CL, only few studies demonstrated higher CL (i.e., task difficulty, mental effort) due to instructor

presence (Homer et al., 2008). Again, some studies did not identify any effects of instructor presence on CL (Colliot & Jamet, 2018; Ng & Przybylek, 2021). Other studies even observed that a visible instructor in videos resulted in lower ratings of extraneous load (Sondermann & Merkt, 2022; Wang, Antonenko, Keil, & Dawson, 2020), intrinsic load (Wang, Antonenko, Keil, & Dawson, 2020), or mental effort (Kizilcec et al., 2015).

Regarding video selection, two studies (Kizilcec at al., 2015; Wilson et al., 2018), including one large-scale observational study in a MOOC environment (Kizilcec at al., 2015), found that learners preferred videos with instructor over those without instructor. Further, an analysis of usage patterns of online educational videos showed that videos with a visible instructor were more engaging because learners watched those videos longer than videos without instructor (Guo et al., 2014).

### 1.3. The present research

There are heterogeneous findings regarding the effects of instructor presence on learning outcomes and subjective ratings of the learning materials (also see Henderson & Schroeder, 2021). This presumably results from investigating the effects of different instructors with various learning materials with different features, which could interact with instructor presence. Hence, we aim to investigate the effects of a talking head across different videos with narrated slides, while systematically varying two additional design features that may moderate effects of instructor presence: First, we include the slide type of the videos (graphic vs. text slides) because based on the split-attention principle (Ayres & Sweller, 2014), it is conceivable that a talking head may provide greater visual competition for graphic-based than for text-based slides, whose visual content is often also covered by the narration. Additionally, the study by Colliot and Jamet (2018) already revealed that a visible instructor had different effects for different types of learning content (e.g., graphic and spoken). Second, we vary the presentation type of the contents in our videos (i.e., whether contents appear sequentially or statically all at once). Since movement attracts attention (Abrams & Christ, 2003), a stepwise fade-in of the slide content may draw learners' attention to the relevant learning contents (cf. signaling principle of the CTML, Mayer & Fiorella, 2014; van Gog, 2014). This may repeatedly direct individuals' attention away from the talking head, thus reducing the talking head's potential distracting effect. We investigate the effects of these different design features on learning outcomes (i.e., factual knowledge) and learners' ratings of the videos (e.g., perceived learning, satisfaction, CL, video selection).

#### 1.3.1. Hypotheses

In a previous study (Sondermann & Merkt, 2022), in which we manipulated the presence of a talking head as a between factor, we found that the combination of talking head, graphic slides, and static presentation was perceived as particularly difficult. For perceived difficulty, we expected to replicate the emerging three-way interaction (Hypothesis 1), which should be reflected in an interaction between slide type and presentation type in the talking head conditions, but not in the conditions without talking head.

Although no learning differences emerged in the previous study (Sondermann & Merkt, 2022), we expected the same interaction of talking head, slide type, and presentation type for learning outcomes (Hypothesis 2) for two reasons: First, because we now used free recall questions instead of multiple-choice questions, which should result in higher difficulty due to the need to retrieve the information instead of just recognising them. Second, for theoretical reasons, because the talking head should be especially hindering to learning when it competes for attention with other visually relevant content, i.e., graphic slides (cf. split-attention principle, Ayres & Sweller, 2014), and when there are no features to direct learners' attention back to the content (cf. signaling principle, van Gog, 2014).

Since previous studies in which learners had a comparison between different formats demonstrated positive effects of instructor presence on subjective ratings (e.g., Wang & Antonenko, 2017; Wang, Antonenko, & Dawson, 2020), we expected higher ratings of perceived learning (Hypothesis 3), satisfaction (Hypothesis 4), and situational interest (Hypothesis 5) for videos with talking head than for videos without talking head. Regarding the selection task, based on the results of two unpublished pilot studies examining the same talking head and a similar slide design, we expected the talking head to have no effect on the selection of the preferred video format (Hypothesis 6). Also based on the pilot studies, we expected an effect of the slide type, with graphic slides being selected more frequently than text slides (Hypothesis 7).

We recorded additional variables (social presence, professionalism) and CL measures (mental effort, extraneous, germane, and intrinsic CL) to identify explanatory mechanisms for potential effects of the talking head on learning outcomes and subjective ratings. Concerning these variables, we preregistered to explore the effects of talking head, slide type, and presentation type due to the heterogeneous findings of prior research (cf. Henderson & Schroeder, 2021). Hypotheses and corresponding analyses were preregistered on OSF (https://osf.io/sd54z/?view_only=43bab4086ffb45989b04f077a925627e).

## 2. Method

### 2.1. Participants and design

We collected 127 complete data sets via Prolific. Based on preregistered exclusion criteria, we excluded 15 participants because they took notes on the videos ($n = 11$), used search engines ($n = 5$), or indicated technical problems ($n = 2$). The excluded cases do not add up to 15 because some participants met multiple exclusion criteria. Exclusions resulted in a final sample size of 112 participants (64 female, 43 male, 5 diverse, $M_{age} = 26.88$ years, $SD_{age} = 9.13$), mostly indicating high school graduation (54%) or a university degree (38%) as their highest level of education. An a priori power analysis with G*Power 3.1.9.4 (Faul, Erdfelder, Lang, & Buchner, 2007) suggested to collect $N = 96$ participants (ANOVA, repeated measures, within factors, $\alpha = 0.05$, $1\text{-}\beta = 0.95$, $f = 0.12$) to detect effects as small as the effects of talking head from previous studies in a within-subjects design. We used a 2x2x2-factorial design with

the three within factors talking head (present vs. absent), slide type (graphic vs. text), and presentation type (sequential vs. static), resulting in eight different video formats.

### 2.2. Materials

The learning materials consisted of eight short videos with narrated slides, each lasting about 1 min and each consisting of one slide. The videos covered factual information about two geographical (Burkina Faso, South Georgia and the South Sandwich Islands), two biological (Ficus lyrata, Blackcap), two physical (Radiocarbon dating, Foucault pendulum), and two historical topics (Wilhelmine Reichard, Rudolf Höß). For each topic, we produced eight content-equivalent formats systematically varying talking head (present vs. absent), slide type (graphic vs. text), and presentation type (sequential vs. static). The narration was identical within a topic across all formats. In videos with talking head, the instructor's head was visible in the lower right corner of the video (see Fig. 1). In videos with sequential presentation, the slide content faded in gradually, whereas in static presentation it appeared all at once. We balanced between participants which video topic they watched in which format. Each participant watched a total of eight videos, one video on each topic and one video in each format. The topics were presented in two orders, whereas the video formats were balanced across positions, which resulted in 16 different sequences (with seven participants per sequence).[1]

### 2.3. Measures

#### 2.3.1. Individual prerequisites

Participants indicated their prior knowledge about the topics covered in the videos on a 7-point Likert scale ranging from 1 (*no prior knowledge*) to 7 (*a lot of prior knowledge*) with one question per topic (e.g., "How much do you know about the African country Burkina Faso?"). In addition, we asked for participants' prior interest in the eight topics (e.g., "How interested are you in the African country Burkina Faso?") on a 7-point Likert scale from 1 (*not at all*) to 7 (*very interested*). Depending on the participants' assignment to sequences, prior knowledge and interest ratings regarding the topics were assigned to the different conditions, to ensure comparability of prior knowledge between conditions independent of topics.

#### 2.3.2. Learning outcomes/knowledge test

We assessed learning outcomes using a self-designed knowledge test with three open recall questions for each of the eight topics (24 items in total). Answers reflected factual information presented on the slides and could be answered from all video versions. Two independent raters assessed the correctness of the responses using a coding scheme. Participants received 1 point for correct, 0.5 points for partially correct, and 0 points for wrong answers. Inter-rater agreement ranged from Cohen's $\kappa = 0.83$ to $\kappa = 1$, indicating almost perfect agreement for all 24 items. Conflicts were resolved by the first author.

#### 2.3.3. Perceived difficulty

After each video, learners rated how easy or difficult the video was to understand on a 7-point scale from 1 (*extremely easy*) to 7 (*extremely difficult*) adapted from Kalyuga, Chandler, and Sweller (1999).

#### 2.3.4. Perceived learning, satisfaction, and situational interest

Learners indicated how much they have learned from the video on a 7-point Likert scale from 1 (*nothing at all*) to 7 (*a lot*) (adapted from Kizilcec et al., 2015; Wang & Antonenko, 2017). Learners rated their satisfaction with the learning materials on a 7-point Likert scale ranging from 1 (*very dissatisfied*) to 7 (*very satisfied*) adapted from Wang and Antonenko (2017). We recorded situational interest with two items (e.g., "Right now I would like to know more about the topic.") adapted from Rotgans and Schmidt (2017) on a 7-point Likert scale ranging from 1 (*strongly disagree*) to 7 (*strongly agree*). The Spearman-Brown coefficient[2] was $\rho = 0.98$.[3]

#### 2.3.5. Selection task

We assessed learners' selection behaviour by informing them at the end of the study that they would now watch another video (topic: sets of numbers). We asked participants to choose one of eight video formats varied before presented as GIFs (without audio track) in randomised order. We informed participants that content and audio track would be identical for all videos and counted how many individuals selected each format.

#### 2.3.6. Social presence

We captured social presence with three items on a 7-point Likert scale from 1 (*strongly disagree*) to 7 (*strongly agree*). Two items (e.g.,

[1] Due to a technical glitch, seven people were assigned to one sequence instead of six, so we decided to fill up all sequences to seven people to ensure perfect balancing which resulted in a total sample size of $N = 112$ participants. As soon as we noticed this glitch during data collection, we noted this modification at OSF in a wiki entry of the project.

[2] For scales with only two items, we used the Spearman-Brown coefficient instead of Cronbach's alpha, because the former is more appropriate for determining reliability of two items (Eisinga, Grotenhuis, & Pelzer, 2013).

[3] Since our within-design provided eight values for each item, we first averaged these eight values and calculated reliabilities of the respective scales based on these mean values.

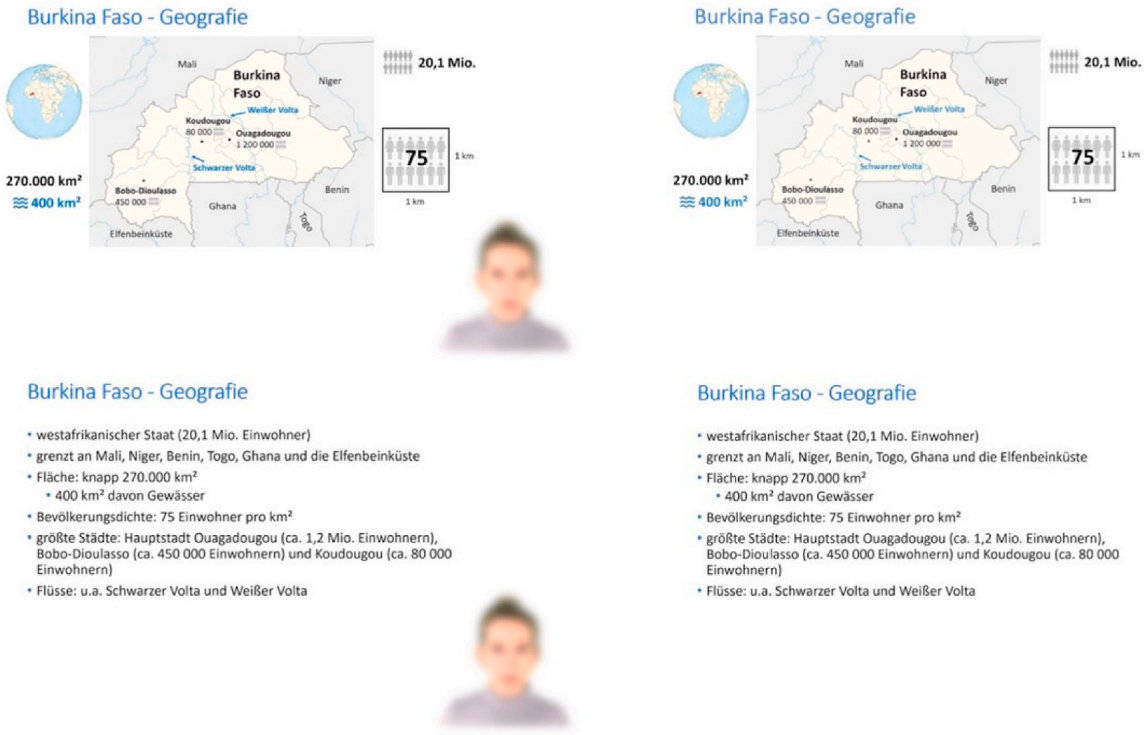



**Fig. 1.** Screenshots of graphic-based (top) and text-based (bottom) videos with (left) and without (right) talking head. The talking head was not blurred in the experiment.

"I felt that the instructor was present.") were taken from a questionnaire used by Kizilcec et al. (2015). We added a third item ("I felt like the instructor addressed me personally.") from the PSI process scales (Schramm & Hartmann, 2008) already applied by Beege, Schneider, Nebel, and Rey (2017). Ratings for the three items were averaged to an overall score (Cronbach's $\alpha = 0.95$).

#### 2.3.7. *Professionalism*

Participants rated the professionalism of the video on a 7-point Likert scale ranging from 1 (*very unprofessional*) to 7 (*very professional*) adapted from Merkt, Lux, Hoogerheide, van Gog, and Schwan (2020).

#### 2.3.8. *Mental effort*

After each video, learners indicated how exhausting it was to follow the video (adapted from Paas, 1992) on a 9-point Likert scale ranging from 1 (*not at all exhausting*) to 9 (*extremely exhausting*).

#### 2.3.9. *Intrinsic, germane, and extraneous CL*

To capture the three different sources of CL, we used the scale of Klepsch, Schmitz, and Seufert (2017). Learners answered eight items on 7-point Likert scales from 1 (*completely wrong*) to 7 (*absolutely right*) with two items addressing intrinsic CL and three items each addressing germane and extraneous CL. Reliabilities were $\rho = 0.92$ for intrinsic, $\alpha = 0.78$ for germane, and $\alpha = 0.89$ for extraneous CL.

### 2.4. *Procedure*

The online experiment was conducted with LimeSurvey. We obtained participants' informed consent and randomly assigned them to one of 16 balancing-sequences. First, participants indicated their prior knowledge and interest about the video contents. After a technical check (video and audio playback), all participants watched eight videos which started automatically and did not allow for user interactions (e.g., pause, replay). After each video, we assessed mental effort, perceived difficulty, situational interest, perceived learning, satisfaction, professionalism, CL, and social presence. After watching all eight videos, participants answered the knowledge test, which was followed by a question on learning preference,[4] demographic questions, and control questions (e.g., use of search engines, note taking behaviour, technical problems). Finally, participants completed the selection task and were debriefed. The

[4] We assessed learning preferences using Verbal-Visual Learning Style Rating (Mayer & Massa, 2003). However, since this concept is critically discussed (e.g., Sinatra & Jacobsen, 2019) and we did not make any assumptions regarding this variable, we decided not to include it in our analyses.

procedure was approved by the local ethics committee.

### 2.5. Analytical strategy

We used repeated measures ANOVAs with the independent variables talking head, slide type, and presentation type. To explain potential interactions, we preregistered to calculate post-hoc tests with Bonferroni-correction. Further, we conducted exploratory mediation analyses to gain insights into explanatory mechanisms for observed effects. Data and scripts can be downloaded at OSF (https://osf.io/afnb8/?view_only=d1b698ba8427488ab3200d36c096fe83).

## 3. Results

A table with the descriptive results of the knowledge test, all rating scales, and the individual prerequisites can be found in the Appendix (Table A.1).

### 3.1. Individual prerequisites

For prior knowledge ($M = 1.29$, $SD = 0.50$), there were neither main effects of talking head, $F(1, 111) = 3.05$, $p = .083$, $\eta_p^2 = 0.027$, slide type, $F < 1$, and presentation type, $F < 1$, nor significant interactions, all $F < 1$. Also for prior interest ($M = 2.73$, $SD = 1.18$), there were neither main effects of talking head, $F(1, 111) = 3.12$, $p = .080$, $\eta_p^2 = 0.027$, slide type, $F < 1$, and presentation type, $F(1, 111) = 1.77$, $p = .186$, $\eta_p^2 = 0.016$, nor significant interactions, all $F < 1.44$, all $p > .233$.

### 3.2. Preregistered confirmatory analyses

#### 3.2.1. Learning outcomes/knowledge test

The ANOVA revealed higher knowledge values for videos without talking head ($M = 1.31$, $SD = 0.55$) compared to videos with talking head ($M = 1.20$, $SD = 0.49$), $F(1, 111) = 4.47$, $p = .037$, $\eta_p^2 = 0.039$. In addition, there were main effects of slide type, $F(1, 111) = 7.81$, $p = .006$, $\eta_p^2 = 0.066$, and presentation type, $F(1, 111) = 4.44$, $p = .037$, $\eta_p^2 = 0.038$. However, there was an interaction Slide Type x Presentation Type, $F(1, 111) = 6.22$, $p = .014$, $\eta_p^2 = 0.053$, which potentially qualifies the latter main effects. Remaining interactions were not significant, all $F < 1$. Resolving the interaction Slide Type x Presentation Type, post-hoc comparisons for graphic-based videos revealed no differences between sequential and static presentation, $F < 1$. For text slides, comparisons indicated higher knowledge scores for static presentation ($M = 1.30$, $SD = 0.71$) compared to sequential presentation ($M = 1.05$, $SD = 0.58$), $F(1, 111) = 11.67$, $p = .001$, $\eta_p^2 = 0.095$.

#### 3.2.2. Perceived difficulty

Learners perceived videos with talking head ($M = 2.08$, $SD = 0.90$) as less difficult than videos without talking head ($M = 2.26$, $SD = 0.84$), $F(1, 111) = 6.46$, $p = .012$, $\eta_p^2 = 0.055$. Further, ANOVA revealed higher difficulty ratings for static ($M = 2.27$, $SD = 0.91$) compared to sequential presentation ($M = 2.08$, $SD = 0.90$), $F(1, 111) = 5.09$, $p = .026$, $\eta_p^2 = 0.044$. Contrary to Hypothesis 1, there was no three-way interaction Instructor Presence x Slide Type x Presentation Type, $F < 1$. There were no other main effects and interactions, all $F < 1.63$, all $p > .204$.

#### 3.2.3. Perceived learning

In line with Hypothesis 3, learners indicated higher perceived learning for videos with talking head ($M = 4.28$, $SD = 0.90$) than for videos without talking head ($M = 4.10$, $SD = 0.98$), $F(1, 111) = 9.56$, $p = .003$, $\eta_p^2 = 0.079$. Further, there was a main effect of slide type, $F(1, 111) = 18.24$, $p < .001$, $\eta_p^2 = 0.141$, which is potentially qualified by the interaction Slide Type x Presentation Type, $F(1, 111) = 6.13$, $p = .015$, $\eta_p^2 = 0.052$. There were no other main effects and interactions, all $F < 1.65$, all $p > .201$. When resolving the interaction Slide Type x Presentation Type, for graphic slides, post-hoc comparisons revealed higher perceived learning for sequential ($M = 4.50$, $SD = 1.09$) than for static presentation ($M = 4.22$, $SD = 1.11$), $F(1, 111) = 7.88$, $p = .006$, $\eta_p^2 = 0.066$. For text slides, there was no difference between sequential and static presentation, $F < 1$.

#### 3.2.4. Satisfaction

In line with Hypothesis 4, videos with talking head ($M = 4.15$, $SD = 1.14$) yielded higher satisfaction ratings than videos without talking head ($M = 3.75$, $SD = 1.17$), $F(1, 111) = 37.30$, $p < .001$, $\eta_p^2 = 0.252$. In addition, the ANOVA revealed main effects of slide type, $F(1, 111) = 32.24$, $p < .001$, $\eta_p^2 = 0.225$, and presentation type, $F(1, 111) = 15.01$, $p < .001$, $\eta_p^2 = 0.119$, as well as an interaction Slide Type x Presentation Type, $F(1, 111) = 9.24$, $p = .003$, $\eta_p^2 = 0.077$. The remaining interactions were not significant, all $F < 1.97$, all $p > .163$. Resolving the interaction Slide Type x Presentation Type, for graphic slides, post-hoc comparisons revealed higher satisfaction for sequential ($M = 4.54$, $SD = 1.28$) than for static presentation ($M = 3.93$, $SD = 1.37$), $F(1, 111) = 23.76$, $p < .001$, $\eta_p^2 = 0.176$. For text slides, there was no difference between sequential and static presentation, $F < 1$.

#### 3.2.5. Situational interest

Contrary to Hypothesis 5, we did not find a main effect of talking head, $F < 1$. The ANOVA only revealed higher situational interest for graphic slides ($M = 3.53$, $SD = 1.08$) than text slides ($M = 3.10$, $SD = 1.14$), $F(1, 111) = 35.67$, $p < .001$, $\eta_p^2 = 0.243$. No main effect

of presentation type, $F(1, 111) = 2.85$, $p = .094$, $\eta_p^2 = 0.025$, and no interactions emerged, all $F < 1.69$, all $p > .196$.

#### 3.2.6. Selection task

Descriptive results of the selection task can be found in the Appendix (Table A.2). Contrary to Hypothesis 6, learners descriptively selected videos with talking head ($M = 0.65$, $SD = 0.48$) more often than videos without talking head ($M = 0.35$, $SD = 0.48$), $F(1,111) = 11.27$, $p = .001$, $\eta_p^2 = 0.092$. In addition, there were main effects of slide type, $F(1,111) = 24.98$, $p < .001$, $\eta_p^2 = 0.184$, and presentation type, $F(1,111) = 64.81$, $p < .001$, $\eta_p^2 = 0.369$, which are potentially qualified by the interaction Slide Type x Presentation Type, $F(1,111) = 16.23$, $p < .001$, $\eta_p^2 = 0.128$. The remaining interactions were not significant, all $F < 1.77$, all $p > .186$. Resolving the interaction Slide Type x Presentation Type, post-hoc comparisons showed that for graphic slides, sequential presentation ($M = 0.60$, $SD = 0.49$) was selected more often than static presentation ($M = 0.12$, $SD = 0.32$), $F(1,111) = 53.55$, $p < .001$, $\eta_p^2 = 0.325$. For text slides, likewise, learners selected sequential presentation ($M = 0.21$, $SD = 0.41$) more often than static presentation ($M = 0.08$, $SD = 0.27$), $F(1,111) = 6.42$, $p = .013$, $\eta_p^2 = 0.055$.

### 3.3. Exploratory analyses

#### 3.3.1. Social presence

Analyses showed higher social presence ratings for videos with talking head ($M = 3.31$, $SD = 1.46$) than for videos without talking head ($M = 2.23$, $SD = 1.18$), $F(1, 111) = 119.18$, $p < .001$, $\eta_p^2 = 0.518$. Further, learners rated social presence higher for graphic slides ($M = 2.81$, $SD = 1.25$) than for text slides ($M = 2.72$, $SD = 1.24$), $F(1, 111) = 4.61$, $p = .034$, $\eta_p^2 = 0.040$, and higher for sequential presentation ($M = 2.85$, $SD = 1.28$) than for static presentation ($M = 2.68$, $SD = 1.24$), $F(1, 111) = 9.12$, $p = .003$, $\eta_p^2 = 0.076$. There were no interactions, all $F < 1$.

#### 3.3.2. Professionalism

There were main effects of talking head, $F(1, 111) = 74.62$, $p < .001$, $\eta_p^2 = 0.402$, slide type, $F(1, 111) = 39.22$, $p < .001$, $\eta_p^2 = 0.261$, and presentation type, $F(1, 111) = 34.66$, $p < .001$, $\eta_p^2 = 0.238$. In addition, analyses showed interactions Talking Head x Slide Type, $F(1, 111) = 7.23$, $p = .008$, $\eta_p^2 = 0.061$, Slide Type x Presentation Type, $F(1, 111) = 12.17$, $p = .001$, $\eta_p^2 = 0.099$, and Talking Head x Slide Type x Presentation Type, $F(1, 111) = 4.35$, $p = .039$, $\eta_p^2 = 0.038$. The interaction Talking Head x Presentation Type was not significant, $F < 1$.

For reasons of space, we only analysed whether the main effect of talking head is qualified by the reported three-way interaction, since we found an effect of talking head on professionalism in a previous study (Sondermann & Merkt, 2022). Post-hoc comparisons indicated no difference in professionalism between videos with and without talking head for graphic slides with sequential presentation, $F(1, 111) = 3.61$, $p = .060$ $\eta_p^2 = 0.032$. However, higher professionalism ratings emerged for videos with talking head than for those without for conditions with graphic slides and static presentation, $F(1, 111) = 21.01$, $p < .001$, $\eta_p^2 = 0.159$, text slides and sequential presentation, $F(1, 111) = 50.01$, $p < .001$, $\eta_p^2 = 0.311$, and text slides and static presentation, $F(1, 111) = 35.92$, $p < .001$, $\eta_p^2 = 0.244$.

#### 3.3.3. Mental effort

Learners indicated higher mental effort for videos without talking head ($M = 3.15$, $SD = 1.25$) than for videos with talking head ($M = 2.85$, $SD = 1.38$), $F(1, 111) = 9.60$, $p = .002$, $\eta_p^2 = 0.080$. Further, analysis revealed higher mental effort for text slides ($M = 3.15$, $SD = 1.57$) than for graphic slides ($M = 2.86$, $SD = 1.18$), $F(1, 111) = 4.87$, $p = .029$, $\eta_p^2 = 0.042$, and for static presentation ($M = 3.17$, $SD = 1.42$) than for sequential presentation ($M = 2.84$, $SD = 1.37$), $F(1, 111) = 6.40$, $p = .013$, $\eta_p^2 = 0.055$. Interactions were not significant, all $F < 1.60$, all $p > .208$.

#### 3.3.4. Intrinsic, germane, and extraneous CL

Regarding intrinsic CL, analyses revealed no main effect of talking head, $F < 1$. However, there were main effects of slide type, $F(1, 111) = 13.28$, $p < .001$, $\eta_p^2 = 0.107$, and presentation type, $F(1, 111) = 4.85$, $p = .030$, $\eta_p^2 = 0.042$. In addition, there was an interaction Slide Type x Presentation Type, $F(1, 111) = 4.29$, $p = .041$, $\eta_p^2 = 0.037$. Other interactions were not significant, all $F < 3.53$, all $p > .062$. Resolving the interaction Slide Type x Presentation Type, for graphic slides, post-hoc comparisons revealed higher intrinsic load values for static ($M = 3.14$, $SD = 1.32$) than for sequential presentation ($M = 2.76$, $SD = 1.23$), $F(1, 111) = 8.32$, $p = .005$, $\eta_p^2 = 0.070$. For text slides, there was no difference between presentation types, $F < 1$.

Concerning germane CL, we found higher ratings for videos with talking head ($M = 4.54$, $SD = 1.02$) than for those without ($M = 4.44$, $SD = 1.11$), $F(1, 111) = 5.39$, $p = .022$, $\eta_p^2 = 0.046$. Further, learners reported higher germane CL for graphic slides ($M = 4.81$, $SD = 1.04$) compared to text slides ($M = 4.17$, $SD = 1.15$), $F(1, 111) = 95.58$, $p < .001$, $\eta_p^2 = 0.463$, and for sequential presentation ($M = 4.55$, $SD = 1.09$) compared to static presentation ($M = 4.43$, $SD = 1.06$), $F(1, 111) = 4.86$, $p = .030$, $\eta_p^2 = 0.042$. There were no interactions, all $F < 3.35$, all $p > .069$.

Regarding extraneous CL, we found lower ratings for videos with talking head ($M = 2.97$, $SD = 0.95$) than for videos without talking head ($M = 3.17$, $SD = 0.98$), $F(1, 111) = 6.45$, $p = .012$, $\eta_p^2 = 0.055$. Further, there were main effects of slide type, $F(1, 111) = 18.53$, $p < .001$, $\eta_p^2 = 0.143$, and presentation type, $F(1, 111) = 19.07$, $p < .001$, $\eta_p^2 = 0.147$, as well as an interaction Slide Type x Presentation Type, $F(1, 111) = 6.19$, $p = .014$, $\eta_p^2 = 0.053$. Remaining interactions were not significant, all $F < 1$. Resolving the interaction Slide Type x Presentation Type, post-hoc comparisons for graphic slides indicated higher extraneous load for static presentation ($M = 3.17$, $SD = 1.34$) than for sequential presentation ($M = 2.48$, $SD = 1.14$), $F(1, 111) = 18.18$, $p < .001$, $\eta_p^2 = 0.141$,

whereas there was no significant difference for text slides, $F(1, 111) = 2.99$, $p = .086$, $\eta_p^2 = 0.026$.

### 3.3.5. Mediation analyses

To explore whether the observed effects of the talking head were mediated by different variables, we conducted mediation analyses using the MEMORE_2.1 macro in SPSS (Montoya & Hayes, 2017), which enables the estimation of mediation models in repeated measures designs. First, based on CTML (Mayer, 2014b), we tested whether the talking head effect on knowledge test performance was mediated by different indicators of CL (i.e., perceived difficulty, mental effort, and extraneous CL). Significant indirect effects were found for perceived difficulty, $b = 0.04$ [CI95%: 0.002, 0.091], mental effort, $b = 0.04$ [CI95%: 0.004, 0.110], and extraneous CL, $b = 0.03$ [CI95%: 0.001, 0.087]. The individual coefficients can be found in Fig. 2. Further, based on SAT (Mayer, 2014a), we tested whether the talking head effect on knowledge test performance was mediated by social presence. Whereas talking head was a significant predictor for social presence (Path a), $b = 1.08$, $t(111) = 10.92$, $p < .001$, social presence was no significant predictor for knowledge test performance when controlling for the effect of the talking head (Path b), $t < 1$, implying that there was no mediation (see Fig. 3). Additionally, based on social-cue hypothesis (Colliot & Jamet, 2018), we aimed to explore whether social presence mediated the talking head effect on satisfaction. The mediation analysis showed a full mediation (see Fig. 3) with a significant indirect effect, $b = 0.31$ [CI95%: 0.175, 0.459].

## 4. Discussion

We investigated the effects of talking heads on learning outcomes, subjective ratings, and selection behaviour, taking into account slide type and presentation type as potential moderators. The experiment provides important insights into the design of educational videos with regard to talking heads. Most importantly, our findings indicate that videos with talking head resulted in poorer learning outcomes than videos without talking head. Regarding self-report measures, however, we observed positive effects of the talking head on perceived learning, satisfaction, social presence, and germane CL. In addition, videos with talking head were preferred in the selection task, and elicited lower ratings of perceived difficulty, mental effort, and extraneous CL, whereas there was no effect on situational interest.

### 4.1. Effects of instructor presence

With regard to learning outcomes, we found no three-way interaction of talking head, slide type, and presentation type (contrary to Hypothesis 2). Instead, videos with talking head led to lower performance in the knowledge test than videos without talking head independent of the other design features. This finding is in line with CLT and CTML (split-attention and coherence principle, Ayres & Sweller, 2014; Mayer & Fiorella, 2014), whereas it does not align with previous research, which mostly showed no effects (e.g.,

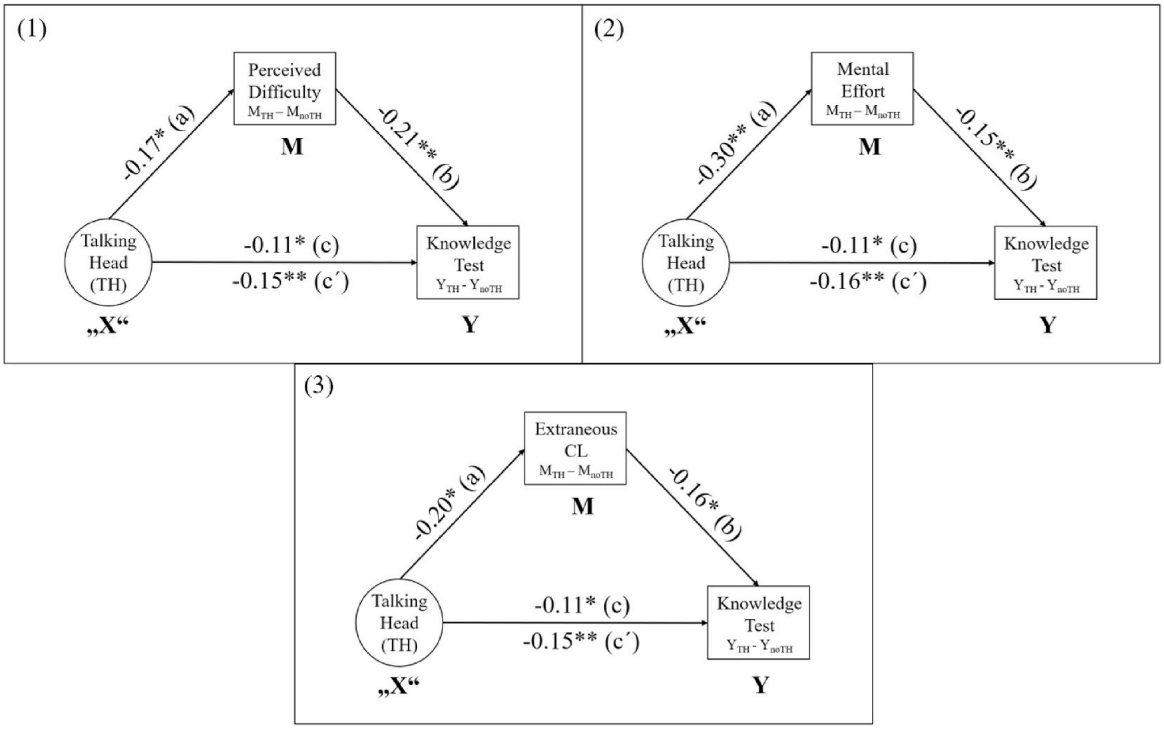



**Fig. 2.** Mediation models showing indirect effects of talking head on knowledge test performance including perceived difficulty (1), mental effort (2), or extraneous CL (3) as mediators.
*p < .05. **p < .01.

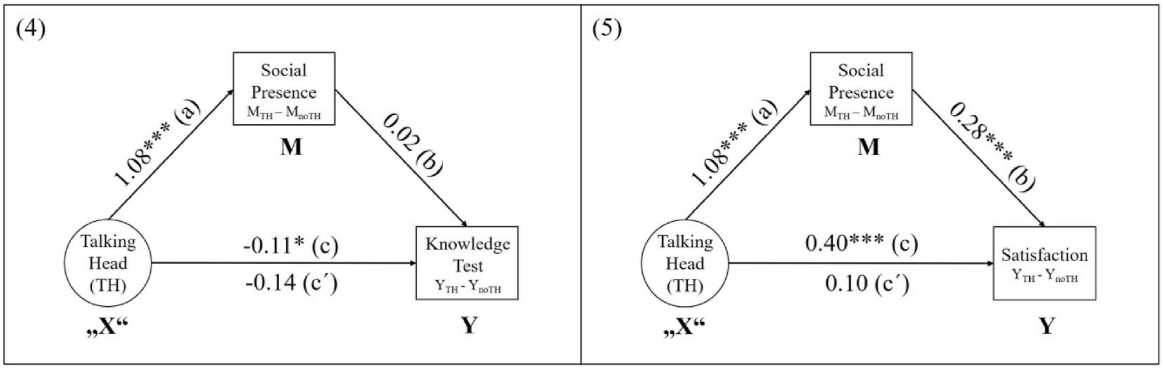



**Fig. 3.** Mediation models showing that social presence did not mediate the talking head effect on knowledge test performance (4), but fully mediated the talking head effect on satisfaction (5).
*$p < .05$. ***$p < .001$.

Kizilcec et al., 2014; van Wermeskerken et al., 2018) or positive effects (e.g., Colliot & Jamet, 2018; Wang & Antonenko, 2017; Wang, Antonenko, & Dawson, 2020) of instructor presence on learning outcomes. A potential explanation for the learning losses might be that the speaking face distracted learners from the actual learning content, as previous research already demonstrated that learners spent a considerable amount of time looking at the instructor (e.g., Kizilcec et al., 2014). In our experiment, learners watched very short videos that conveyed a lot of information in a very short time. Hence, even small distractions could lead to learning losses. It is conceivable that in previous studies that did not identify negative effects (e.g., Kizilcec et al., 2014), the longer videos may have been characterised by reduced information density, so that learners were able to compensate for short distractions caused by the instructor.

Although our experiment showed a negative effect of the talking head on learning, we could not identify moderators of this instructor presence effect, contrary to our expectations. One possible explanation for the missing moderation might include that the design characteristics which we additionally varied as independent variables (e.g., sequential fade-in of new content) were too weak to modulate the attention-drawing effects of the talking head in our experiment. However, it should be noted that there were interactions between slide type and presentation type independent of the talking head manipulation, so that it would not be appropriate to state that other video characteristics did not have an effect at all. Additionally, it is conceivable that potential differences between the conditions did not manifest in the measures we collected, as they were all recorded after the respective learning phase. In particular, it cannot be ruled out that there were different processes by which learners in the different conditions achieved the same result. This explanation could be validated, for example, by analysing participants' gaze behaviour to investigate whether there were specific patterns for conditions with and without talking head depending on slide type and/or presentation type.

Contrary to Hypothesis 1, there was no three-way interaction for perceived difficulty. Instead, participants perceived videos without talking head as more difficult than videos with talking head across different conditions. Correspondingly, they expressed lower mental effort and extraneous load for videos with talking head. These findings do not align with our findings on learning outcomes and theory. According to the coherence and split-attention principle, irrelevant elements in the learning material as well as split-attention caused by the instructor's video should increase extraneous CL (Ayres & Sweller, 2014; Mayer & Fiorella, 2014). However, consistent with our findings, previous studies also reported lower extraneous load (Sondermann & Merkt, 2022; Wang, Antonenko, Keil, & Dawson, 2020) and mental effort (Kizilcec et al., 2015) for videos including an instructor. It is conceivable that lower ratings on various CL measures result from participants often looking at the face (Kizilcec et al., 2014) and presumably paying less attention to the actual learning content. Therefore, learners might be less able to assess its characteristics (e.g., its difficulty), whereas they are unlikely to perceive the instructor's face itself as particularly difficult or cognitively demanding. Besides, the exploratory mediation analyses provided a deeper insight into the explanatory mechanisms for the instructor presence effect. These analyses showed indirect effects of the talking head on learning outcomes when including different indicators of CL (i.e., perceived difficulty, mental effort, extraneous CL) as mediators. The easier (or less effortful or less strenuous) the learners rated the videos with talking head compared to videos without talking head, the less hindering to learning the talking head was compared to videos without talking head.

Corroborating Hypothesis 3, the talking head resulted in increased ratings of perceived learning in line with prior research (e.g., Kizilcec et al., 2015). When comparing the findings on perceived learning with those from the knowledge test, participants perceived to have learned better with videos including a talking head, whereas they actually performed worse with these videos. Hence, in this experiment, the talking head induced a knowledge illusion.

Our results further provide evidence for Hypothesis 4 in that the talking head resulted in higher satisfaction ratings. These findings are in line with previous research in which participants had a comparison between formats with and without instructor (e.g., Wang & Antonenko, 2017). Results of our mediation analyses revealed that the higher ratings of satisfaction are mediated by higher ratings of social presence. Hence, our findings are also consistent with social-cue hypothesis (Colliot & Jamet, 2018), suggesting that social cues in the learning material positively affect aspects such as learners' engagement and motivation.

Further, in contrast to Hypothesis 6, but in line with the results of subjective rating scales, participants selected videos with talking head more frequently than videos without talking head, which aligns with previous research (Kizilcec et al., 2015; Wilson et al., 2018). However, contrary to Hypothesis 5, the talking head did not affect participants' situational interest. This finding is consistent with the results of Colliot and Jamet (2018), but it does not fit with Wang, Antonenko, and Dawson (2020) and the social-cue hypothesis

(Colliot & Jamet, 2018). This missing effect cannot be explained by the talking head being too weak as a social cue, because there was a particularly large effect of the talking head on social presence.

Further, videos with talking head elicited higher ratings of germane CL, which is in line with Mayer's SAT (2014a) postulating that social cues (e.g., a visible instructor) can trigger social responses, that in turn lead to deeper processing. Contrary to SAT, we did not observe beneficial effects on learning outcomes resulting from deeper processing, but rather found detrimental effects on learning outcomes. However, it should be noted that learning-enhancing effects of deeper processing according to SAT (Mayer, 2014a) should be more apparent in higher order learning measures (i.e., transfer) and not in factual knowledge acquisition which was the scope of the current experiment.

### 4.2. Effects of slide type and presentation type

In contrast to Hypothesis 7, learners did not generally select graphic slides more frequently than text slides. Such a preference was only found for videos with sequential presentation. However, graphic slides increased situational interest compared to text slides. It is conceivable that the numerous images and illustrations on graphic slides increased interest. In addition, graphic slides led to higher social presence ratings, which may result from pictures (e.g., of people) that were part of the learning material. In addition, learners experienced less mental effort with graphic slides than with text slides, which could result from the fact that the information on graphic slides was presented via two codes (i.e., dual coding, Paivio, 1990).

Regarding presentation type, learners rated perceived difficulty and mental effort lower for videos with sequential presentation compared to those with static presentation. These ratings fit the signaling principle (Mayer & Fiorella, 2014; van Gog, 2014), as sequential presentation provides learners with cues about where to shift their attention and how to organise the material, thus reducing the processing of extraneous material. However, these ratings are not reflected in the knowledge test results, as participants did not show better learning outcomes with sequentially presented content, in line with the results of a prior study (Sondermann & Merkt, 2022). Further, learners preferred sequential presentation over static presentation in the selection task and expressed higher social presence ratings for sequential presentation. Possibly, this effect may be explained by participants' attributions of the sequential presentation to an instructor advancing through the slides.

Our results revealed a consistent interaction of slide type and presentation type for several measures (e.g., perceived learning, satisfaction, extraneous load). Videos with graphic slides induced higher ratings of perceived learning and satisfaction and lower ratings of extraneous and intrinsic load with sequentially presented content than with statically presented content. For text slides, there were no differences between static and sequential presentation on any of the measures mentioned above. This pattern of findings could be explained by the fact that in videos with graphic slides, order and organisation of the presented content were not obvious to the participants. Therefore, sequential presentation potentially guided them where to shift their attention by fading the content in step by step (cf. signaling principle, van Gog, 2014). The observation that there was no advantage of sequential presentation over static presentation for text slides may be explained by the clear reading direction of text slides (from left to right and from top to bottom). Thus, sequential presentation in text slides possibly did not offer a crucial advantage.

### 4.3. Limitations and directions for future research

The aforementioned findings should be interpreted in the light of some methodological choices: First, our knowledge test measured free recall of factual information and did not include higher order learning (i.e., transfer knowledge). Hence, our conclusions regarding learning outcomes refer to acquisition of factual knowledge and cannot be generalised to other learning measures. However, we focused on factual knowledge because a potential distraction by the talking head should affect learners' extraction of factual information from the learning materials (i.e., when learners do not attend to the actual slide contents but focus on the head instead). This extraction of factual information may also be important in more complex learning tasks, as it provides the prerequisites for deeper understanding (see CTML, Mayer, 2014b).

Second, it is important to note that we recorded learning outcomes in the knowledge test with only three items per topic, although it would certainly have been beneficial to include even more questions per topic (e.g., to enable more variance in the answers). However, due to the short length of the videos, it was not possible to develop more than three appropriate questions. Nevertheless, it is important to consider that this limited number of questions per topic still produced sufficient variability to detect differences between the conditions, even though the observed effect was admittedly small.

Third, it should be noted that we decided to use those rather short videos (about 1 min each) in order to examine videos on different topics. However, this methodological choice also implies that our findings cannot be generalised to all types of learning videos (e.g., videos of a 90-min lecture) without further research.

Further research is also needed regarding the observed learning-impeding effect of the instructor, because we were not able to capture participants' eye movements in this online experiment. To understand why participants' learning performance decreased with videos including a talking head, even though they rated these videos as more positive and less challenging, future studies should use eye-tracking to determine learners' gaze behaviour. This would allow inferences on whether poorer learning outcomes are associated with increased visual attention on the talking head and whether there are associations between attention on the talking head and various CL measures.

### 4.4. Implications and conclusion

This experiment offers numerous implications for the design of educational videos. Inserting the instructor resulted in better ratings of the videos. Learners reported higher satisfaction, better perceived learning, higher social presence, and lower CL, and were more likely to select these videos. Hence, according to our results, designers of educational videos who aim to increase the popularity of their videos can achieve this by integrating a talking head. However, it should be kept in mind that the integration of the talking head may come at the expense of learning outcomes. More specifically, the results of our knowledge test suggest that the visible instructor can have a small detrimental effect on learning. Accordingly, the decision whether to include a talking head in educational videos may involve a trade-off between desired popularity and educational use, while it is also important not to neglect indirect effects of subjective ratings on learning outcomes. For example, in free-choice learning situations, positive ratings may be a precondition for watching videos in the first place. Our findings also imply that when learners are given a choice between video formats, they do not necessarily choose formats with which they learn best. Thus, participants in our study selected videos with instructor more often, even though learning was worse with them. Further, the interplay of slide type and presentation type observed consistently across different measures (e.g., satisfaction, perceived learning), suggests that the combination of graphic slides and sequential presentation leads to particularly positive results. Hence, this format can be recommended for the design of educational videos.

## Author statement

**Christina Sondermann:** Writing- Original draft, Writing- Reviewing and Editing, Conceptualization, Methodology, Project administration, Investigation, Formal analysis, Data curation; **Martin Merkt:** Writing- Reviewing and Editing, Conceptualization, Methodology, Project administration. Supervision.

## Declaration of competing interest

None.

## Data availability

The OSF-link to data/code is included in the manuscript.

## Appendix

**Table A.1**
Means (and Standard Deviations in Parentheses) for the Knowledge Test and Rating Scales

| Condition | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Instructor | Slide type | Presentation type | KT | PD | PL | SAT | SI | SP | PRO | ME | ICL | GCL | ECL | PK | PI |
| Present | Graphic | Sequential | 1.30 (0.95) | 1.88 (1.11) | 4.60 (1.25) | 4.73 (1.52) | 3.63 (1.67) | 3.44 (1.55) | 4.52 (1.46) | 2.46 (1.72) | 2.60 (1.42) | 4.91 (1.24) | 2.37 (1.31) | 1.33 (0.95) | 2.63 (1.55) |
| | | Static | 1.26 (0.82) | 2.23 (1.37) | 4.33 (1.30) | 4.04 (1.58) | 3.33 (1.70) | 3.24 (1.55) | 3.96 (1.51) | 3.02 (1.99) | 3.21 (1.62) | 4.75 (1.15) | 3.11 (1.58) | 1.24 (0.74) | 2.71 (1.63) |
| | | Overall | 1.28 (0.64) | 2.06 (1.02) | 4.46 (1.01) | 4.39 (1.28) | 3.48 (1.30) | 3.34 (1.48) | 4.24 (1.33) | 2.74 (1.48) | 2.91 (1.26) | 4.83 (1.08) | 2.74 (1.07) | 1.29 (0.74) | 2.67 (1.38) |
| | Text | Sequential | 1.03 (0.79) | 2.01 (1.16) | 4.07 (1.34) | 4.00 (1.58) | 3.26 (1.64) | 3.35 (1.55) | 4.04 (1.46) | 2.83 (1.98) | 2.65 (1.62) | 4.30 (1.29) | 3.08 (1.57) | 1.21 (0.64) | 2.65 (1.69) |
| | | Static | 1.20 (0.88) | 2.21 (1.37) | 4.13 (1.37) | 3.84 (1.57) | 3.07 (1.60) | 3.20 (1.56) | 3.75 (1.42) | 3.10 (2.25) | 2.70 (1.44) | 4.21 (1.24) | 3.33 (1.50) | 1.21 (0.69) | 2.60 (1.66) |
| | | Overall | 1.12 (0.66) | 2.11 (1.03) | 4.10 (1.09) | 3.92 (1.30) | 3.17 (1.29) | 3.27 (1.50) | 3.89 (1.27) | 2.96 (1.75) | 2.67 (1.34) | 4.25 (1.15) | 3.20 (1.31) | 1.21 (0.51) | 2.63 (1.46) |
| | Overall | Sequential | 1.17 (0.64) | 1.95 (0.94) | 4.33 (1.03) | 4.37 (1.29) | 3.45 (1.39) | 3.39 (1.50) | 4.28 (1.26) | 2.64 (1.49) | 2.63 (1.33) | 4.61 (1.13) | 2.72 (1.16) | 1.27 (0.63) | 2.64 (1.43) |
| | | Static | 1.23 (0.62) | 2.22 (1.12) | 4.23 (1.12) | 3.94 (1.32) | 3.20 (1.38) | 3.22 (1.50) | 3.86 (1.27) | 3.06 (1.78) | 3.00 (1.23) | 4.48 (1.04) | 3.22 (1.14) | 1.22 (0.64) | 2.65 (1.44) |
| | | Overall | 1.20 (0.49) | 2.08 (0.90) | 4.28 (0.90) | 4.15 (1.14) | 3.33 (1.12) | 3.31 (1.46) | 4.07 (1.19) | 2.85 (1.38) | 2.79 (1.14) | 4.54 (1.02) | 2.97 (0.95) | 1.25 (0.56) | 2.65 (1.29) |
| Absent | Graphic | Sequential | 1.39 (0.87) | 2.13 (1.36) | 4.39 (1.34) | 4.34 (1.55) | 3.71 (1.63) | 2.38 (1.45) | 4.27 (1.43) | 2.77 (1.96) | 2.91 (1.50) | 4.92 (1.16) | 2.59 (1.46) | 1.35 (0.85) | 2.73 (1.52) |
| | | Static | 1.39 (0.92) | 2.30 (1.20) | 4.11 (1.33) | 3.82 (1.60) | 3.44 (1.67) | 2.18 (1.24) | 3.38 (1.51) | 3.20 (2.04) | 3.06 (1.53) | 4.66 (1.27) | 3.23 (1.69) | 1.32 (0.83) | 2.92 (1.77) |
| | | Overall | 1.39 (0.70) | 2.22 (0.99) | 4.25 (1.15) | 4.08 (1.27) | 3.57 (1.23) | 2.28 (1.25) | 3.83 (1.24) | 2.98 (1.39) | 2.99 (1.23) | 4.79 (1.12) | 2.91 (1.13) | 1.33 (0.64) | 2.83 (1.41) |

(continued on next page)

**Table A.1** (*continued*)

| Condition | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Instructor | Slide type | Presentation type | KT | PD | PL | SAT | SI | SP | PRO | ME | ICL | GCL | ECL | PK | PI |
| | Text | Sequential | 1.08 | 2.29 | 3.90 | 3.42 | 2.96 | 2.23 | 3.20 | 3.29 | 2.61 | 4.06 | 3.34 | 1.31 | 2.66 |
| | | | (0.79) | (1.45) | (1.27) | (1.60) | (1.57) | (1.35) | (1.56) | (2.14) | (1.61) | (1.39) | (1.65) | (0.85) | (1.57) |
| | | Static | 1.40 | 2.31 | 4.02 | 3.42 | 3.10 | 2.11 | 3.09 | 3.36 | 2.59 | 4.12 | 3.50 | 1.35 | 2.91 |
| | | | (0.94) | (1.32) | (1.37) | (1.66) | (1.80) | (1.30) | (1.43) | (2.14) | (1.44) | (1.34) | (1.62) | (0.95) | (1.68) |
| | | Overall | 1.24 | 2.30 | 3.96 | 3.42 | 3.03 | 2.17 | 3.14 | 3.33 | 2.60 | 4.09 | 3.42 | 1.33 | 2.79 |
| | | | (0.67) | (1.08) | (1.07) | (1.42) | (1.35) | (1.19) | (1.33) | (1.67) | (1.27) | (1.27) | (1.31) | (0.70) | (1.38) |
| | Overall | Sequential | 1.23 | 2.21 | 4.15 | 3.88 | 3.33 | 2.31 | 3.73 | 3.03 | 2.76 | 4.49 | 2.97 | 1.33 | 2.70 |
| | | | (0.64) | (1.13) | (1.11) | (1.27) | (1.34) | (1.29) | (1.28) | (1.63) | (1.32) | (1.13) | (1.23) | (0.71) | (1.32) |
| | | Static | 1.39 | 2.31 | 4.06 | 3.62 | 3.27 | 2.15 | 3.24 | 3.28 | 2.83 | 4.39 | 3.37 | 1.33 | 2.92 |
| | | | (0.71) | (0.97) | (1.23) | (1.37) | (1.43) | (1.19) | (1.27) | (1.56) | (1.21) | (1.20) | (1.26) | (0.67) | (1.51) |
| | | Overall | 1.31 | 2.26 | 4.10 | 3.75 | 3.30 | 2.23 | 3.48 | 3.15 | 2.79 | 4.44 | 3.17 | 1.33 | 2.81 |
| | | | (0.55) | (0.84) | (0.98) | (1.17) | (1.14) | (1.18) | (1.15) | (1.25) | (1.10) | (1.11) | (0.98) | (0.56) | (1.25) |
| Overall | Graphic | Sequential | 1.34 | 2.01 | 4.50 | 4.54 | 3.67 | 2.91 | 4.39 | 2.61 | 2.76 | 4.92 | 2.48 | 1.34 | 2.68 |
| | | | (0.69) | (0.97) | (1.09) | (1.28) | (1.26) | (1.35) | (1.27) | (1.53) | (1.23) | (1.09) | (1.14) | (0.71) | (1.25) |
| | | Static | 1.33 | 2.27 | 4.22 | 3.93 | 3.39 | 2.71 | 3.67 | 3.11 | 3.14 | 4.71 | 3.17 | 1.28 | 2.81 |
| | | | (0.67) | (1.08) | (1.11) | (1.37) | (1.35) | (1.26) | (1.36) | (1.56) | (1.32) | (1.11) | (1.34) | (0.65) | (1.46) |
| | | Overall | 1.33 | 2.14 | 4.36 | 4.23 | 3.53 | 2.81 | 4.03 | 2.86 | 2.95 | 4.81 | 2.82 | 1.31 | 2.75 |
| | | | (0.54) | (0.84) | (0.97) | (1.15) | (1.08) | (1.25) | (1.17) | (1.18) | (1.07) | (1.04) | (0.90) | (0.59) | (1.20) |
| | Text | Sequential | 1.05 | 2.15 | 3.99 | 3.71 | 3.11 | 2.79 | 3.62 | 3.06 | 2.63 | 4.18 | 3.21 | 1.26 | 2.66 |
| | | | (0.58) | (1.12) | (1.13) | (1.45) | (1.29) | (1.31) | (1.38) | (1.74) | (1.40) | (1.26) | (1.37) | (0.57) | (1.35) |
| | | Static | 1.30 | 2.26 | 4.07 | 3.63 | 3.09 | 2.66 | 3.42 | 3.23 | 2.65 | 4.16 | 3.41 | 1.28 | 2.75 |
| | | | (0.71) | (1.13) | (1.16) | (1.43) | (1.38) | (1.29) | (1.30) | (1.95) | (1.31) | (1.18) | (1.32) | (0.65) | (1.43) |
| | | Overall | 1.18 | 2.21 | 4.03 | 3.67 | 3.10 | 2.72 | 3.52 | 3.15 | 2.64 | 4.17 | 3.31 | 1.27 | 2.71 |
| | | | (0.52) | (0.93) | (0.98) | (1.29) | (1.14) | (1.24) | (1.22) | (1.57) | (1.20) | (1.15) | (1.19) | (0.50) | (1.23) |
| | Overall | Sequential | 1.20 | 2.08 | 4.24 | 4.12 | 3.39 | 2.85 | 4.00 | 2.84 | 2.69 | 4.55 | 2.84 | 1.30 | 2.67 |
| | | | (0.51) | (0.90) | (0.96) | (1.17) | (1.11) | (1.28) | (1.19) | (1.37) | (1.19) | (1.09) | (1.03) | (0.56) | (1.19) |
| | | Static | 1.31 | 2.27 | 4.15 | 3.78 | 3.24 | 2.68 | 3.55 | 3.17 | 2.89 | 4.43 | 3.29 | 1.28 | 2.78 |
| | | | (0.54) | (0.91) | (0.98) | (1.22) | (1.18) | (1.24) | (1.19) | (1.42) | (1.11) | (1.06) | (1.02) | (0.55) | (1.33) |
| | | Overall | 1.26 | 2.17 | 4.19 | 3.95 | 3.31 | 2.77 | 3.78 | 3.00 | 2.79 | 4.49 | 3.07 | 1.29 | 2.73 |
| | | | (0.44) | (0.79) | (0.89) | (1.10) | (1.04) | (1.22) | (1.11) | (1.21) | (1.04) | (1.04) | (0.87) | (0.50) | (1.18) |

*Note.* KT = knowledge test; PD = perceived difficulty; PL = perceived learning; SAT = satisfaction; SI = situational interest; SP = social presence; PRO = professionalism; ME = mental effort; ICL = intrinsic cognitive load; GCL = germane cognitive load; ECL = extraneous cognitive load; PK = prior knowledge; PI = prior interest.

**Table A.2**
Means and Standard Deviations of the Selection Frequencies for the Selection Task

| Talking head | Slide type | Presentation type | *M* | *SD* |
|---|---|---|---|---|
| Present | Graphic | Sequential | 0.37 | 0.48 |
| | | Static | 0.09 | 0.29 |
| | | Overall | 0.46 | 0.50 |
| | Text | Sequential | 0.14 | 0.35 |
| | | Static | 0.05 | 0.23 |
| | | Overall | 0.20 | 0.40 |
| | Overall | Sequential | 0.51 | 0.50 |
| | | Static | 0.14 | 0.35 |
| | | Overall | 0.65 | 0.48 |
| Absent | Graphic | Sequential | 0.23 | 0.42 |
| | | Static | 0.03 | 0.16 |
| | | Overall | 0.26 | 0.44 |
| | Text | Sequential | 0.06 | 0.24 |
| | | Static | 0.03 | 0.16 |
| | | Overall | 0.09 | 0.29 |
| | Overall | Sequential | 0.29 | 0.46 |
| | | Static | 0.05 | 0.23 |
| | | Overall | 0.35 | 0.48 |
| Overall | Graphic | Sequential | 0.60 | 0.49 |
| | | Static | 0.12 | 0.32 |
| | | Overall | 0.71 | 0.45 |
| | Text | Sequential | 0.21 | 0.41 |
| | | Static | 0.08 | 0.27 |
| | | Overall | 0.29 | 0.45 |
| | Overall | Sequential | 0.80 | 0.40 |
| | | Static | 0.20 | 0.40 |
| | | Overall | 1.00 | 0.00 |

## References


Abrams, R. A., & Christ, S. E. (2003). Motion onset captures attention. *Psychological Science, 14*, 427–432. https://doi.org/10.1111/1467-9280.01458

Ayres, P., & Sweller, J. (2014). The split-attention principle in multimedia learning. In R. E. Mayer (Ed.), *The Cambridge handbook of multimedia learning* (2nd ed., pp. 206–226). Cambridge University Press.

Beege, M., Ninaus, M., Schneider, S., Nebel, S., Schlemmel, J., Weidenmüller, J., et al. (2020). Investigating the effects of beat and deictic gestures of a lecturer in educational videos. *Computers & Education, 156*, Article 103955. https://doi.org/10.1016/j.compedu.2020.103955

Beege, M., Schneider, S., Nebel, S., & Rey, G. D. (2017). Look into my eyes! Exploring the effect of addressing in educational videos. *Learning and Instruction, 49*, 113–120. https://doi.org/10.1016/j.learninstruc.2017.01.004

Colliot, T., & Jamet, É. (2018). Understanding the effects of a teacher video on learning from a multimedia document: An eye-tracking study. *Educational Technology Research & Development, 66*, 1415–1433. https://doi.org/10.1007/s11423-018-9594-x

Eisinga, R., Grotenhuis, M.t., & Pelzer, B. (2013). The reliability of a two-item scale: Pearson, Cronbach, or Spearman-Brown? *International Journal of Public Health, 58*, 637–642. https://doi.org/10.1007/s00038-012-0416-3

Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G*power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods, 39*, 175–191. https://doi.org/10.3758/BF03193146

Fiorella, L., Stull, A. T., Kuhlmann, S., & Mayer, R. E. (2019). Instructor presence in video lectures: The role of dynamic drawings, eye contact, and instructor visibility. *Journal of Educational Psychology, 111*, 1162–1171. https://doi.org/10.1037/edu0000325

van Gog, T. (2014). The signaling (or cueing) principle in multimedia learning. In R. E. Mayer (Ed.), *The Cambridge handbook of multimedia learning* (2nd ed., pp. 263–278). Cambridge University Press.

Gullberg, M., & Holmqvist, K. (2006). What speakers do and what addressees look at. *Pragmatics and Cognition, 14*, 53–82. https://doi.org/10.1075/pc.14.1.05gul

Guo, P. J., Kim, J., & Rubin, R. (2014). How video production affects student engagement. In M. Sahami, A. Fox, M. A. Hearst, & M. T. Chi (Eds.), *Proceedings of the first ACM conference on Learning @ scale conference - l@S '14* (pp. 41–50). ACM Press. https://doi.org/10.1145/2556325.2566239.

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

Homer, B. D., Plass, J. L., & Blake, L. (2008). The effects of video on cognitive load and social presence in multimedia-learning. *Computers in Human Behavior, 24*, 786–797. https://doi.org/10.1016/j.chb.2007.02.009

Hong, J., Pi, Z., & Yang, J. (2018). Learning declarative and procedural knowledge via video lectures: Cognitive load and learning effectiveness. *Innovations in Education & Teaching International, 55*, 74–81. https://doi.org/10.1080/14703297.2016.1237371

Johnson, M. H., Dziurawiec, S., Ellis, H., & Morton, J. (1991). Newborns' preferential tracking of face-like stimuli and its subsequent decline. *Cognition, 40*, 1–19. https://doi.org/10.1016/0010-0277(91)90045-6

Kalyuga, S., Chandler, P., & Sweller, J. (1999). Managing split-attention and redundancy in multimedia instruction. *Applied Cognitive Psychology, 13*, 351–371. https://doi.org/10.1002/(SICI)1099-0720 (199908)13:4<351::AID-ACP589>3.0.CO;2-6.

Kizilcec, R. F., Bailenson, J. N., & Gomez, C. J. (2015). The instructor's face in video instruction: Evidence from two large-scale field studies. *Journal of Educational Psychology, 107*, 724–739. https://doi.org/10.1037/edu0000013

Kizilcec, R. F., Papadopoulos, K., & Sritanyaratana, L. (2014). Showing face in video instruction: Effects on information retention, visual attention, and affect. In M. Jones, P. Palanque, A. Schmidt, & T. Grossman (Eds.), *Proceedings of the 32nd annual ACM conference on Human factors in computing systems - CHI '14* (pp. 2095–2102). https://doi.org/10.1145/2556288.2557207. Association for Computing Machinery.

Klepsch, M., Schmitz, F., & Seufert, T. (2017). Development and validation of two instruments measuring intrinsic, extraneous, and germane cognitive load. *Frontiers in Psychology, 8*. https://doi.org/10.3389/fpsyg.2017.01997, 1997.

Mayer, R. E. (2014a). Principles based on social cues in multimedia learning: Personalization, voice, image and embodiment principles. In R. E. Mayer (Ed.), *The Cambridge handbook of multimedia learning* (2nd ed., pp. 345–368). Cambridge University Press.

Mayer, R. E. (Ed.). (2014b). *The Cambridge handbook of multimedia learning* (2nd ed.). Cambridge University Press. https://doi.org/10.1017/CBO9781139547369.

Mayer, R. E., & Fiorella, L. (2014). Principles for reducing extraneous processing in multimedia learning: Coherence, signaling, redundancy, spatial contiguity, and temporal contiguity principles. In R. E. Mayer (Ed.), *The Cambridge handbook of multimedia learning* (2nd ed., pp. 279–315). Cambridge University Press.

Mayer, R. E., & Massa, L. J. (2003). Three facets of visual and verbal learners: Cognitive ability, cognitive style, and learning Preference. *Journal of Educational Psychology, 95*, 833–846. https://doi.org/10.1037/0022-0663.95.4.833

Merkt, M., Lux, S., Hoogerheide, V., van Gog, T., & Schwan, S. (2020). A change of scenery: Does the setting of an instructional video affect learning? *Journal of Educational Psychology, 112*, 1273–1283. https://doi.org/10.1037/edu0000414

Montoya, A. K., & Hayes, A. F. (2017). Two-condition within-participant statistical mediation analysis: A path-analytic framework. *Psychological Methods, 22*, 6–27. https://doi.org/10.1037/met0000086

Ng, Y. Y., & Przybylek, A. (2021). Instructor presence in video lectures: Preliminary findings from an online experiment. *IEEE Access, 9*, 36485–36499. https://doi.org/10.1109/ACCESS.2021.3058735

Paas, F. G. (1992). Training strategies for attaining transfer of problem-solving skill in statistics: A cognitive-load approach. *Journal of Educational Psychology, 84*, 429–434. https://doi.org/10.1037/0022-0663.84.4.429

Paivio, A. (1990). *Mental representations*. Oxford University Press. https://doi.org/10.1093/acprof:oso/9780195066661.001.0001.

Pi, Z., & Hong, J. (2016). Learning process and learning outcomes of video podcasts including the instructor and PPT slides: A Chinese case. *Innovations in Education & Teaching International, 53*, 135–144. https://doi.org/10.1080/14703297.2015.1060133

Rotgans, J. I., & Schmidt, H. G. (2017). Interest development: Arousing situational interest affects the growth trajectory of individual interest. *Contemporary Educational Psychology, 49*, 175–184. https://doi.org/10.1016/j.cedpsych.2017.02.003

Schneider, S., Beege, M., Nebel, S., Schnaubert, L., & Rey, G. D. (2022). The cognitive-affective-social theory of learning in digital environments (CASTLE). *Educational Psychology Review, 34*, 1–38. https://doi.org/10.1007/s10648-021-09626-5

Schramm, H., & Hartmann, T. (2008). The PSI-Process Scales. A new measure to assess the intensity and breadth of parasocial processes. *Communications, 33*, 385–401. https://doi.org/10.1515/COMM.2008.025

Sinatra, G. M., & Jacobsen, N. G. (2019). Zombie concepts in education: Why they won't die and why you can't kill them. In P. Kenedeou, D. H. Robinson, & M. T. McCrudden (Eds.), *Misinformation and fake news in education* (pp. 7–27). Information Age Publishing.

Sondermann, C., & Merkt, M. (2022). *What is the effect of talking heads in educational videos with different types of narrated slides? [Manuscript submitted for publication]*.

Stull, A. T., Fiorella, L., & Mayer, R. E. (2021). The case for embodied instruction: The instructor as a source of attentional and social cues in video lectures. *Journal of Educational Psychology, 113*, 1441–1453. https://doi.org/10.1037/edu0000650

Sweller, J. (1999). *Instructional design in technical areas*. ACER Press.

Sweller, J., Ayres, P., & Kalyuga, S. (2011). *Cognitive load theory*. Springer.

Wang, J., & Antonenko, P. (2017). Instructor presence in instructional video: Effects on visual attention, recall, and perceived learning. *Computers in Human Behavior, 71*, 79–89. https://doi.org/10.1016/j.chb.2017.01.049

Wang, J., Antonenko, P., & Dawson, K. (2020). Does visual attention to the instructor in online video affect learning and learner perceptions? An eye-tracking analysis. *Computers & Education, 146*, Article 103779. https://doi.org/10.1016/j.compedu.2019.103779

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*, 279–291. https://doi.org/10.1111/mbe.12239

van Wermeskerken, M., Ravensbergen, S., & van Gog, T. (2018). Effects of instructor presence in video modeling examples on attention and learning. *Computers in Human Behavior, 89*, 430–438. https://doi.org/10.1016/j.chb.2017.11.038

Wilson, K. E., Martinez, M., Mills, C., D'Mello, S., Smilek, D., & Risko, E. F. (2018). Instructor presence effect: Liking does not always lead to learning. *Computers & Education, 122*, 205–220. https://doi.org/10.1016/j.compedu.2018.03.011

# Appendix C

Manuscript 3

Sondermann, C., Huff, M., & Merkt, M. (2024). Distracted by a talking head? An eye tracking study on the effects of instructor presence in learning videos with animated graphic slides. *Learning and Instruction, 91*, Article 101878. https://doi.org/10.1016/j.learninstruc.2024.101878

| **Author** | **Author position** | **Scientific ideas %** | **Data generation %** | **Analysis and interpretation %** | **Paper writing %** |
|---|---|---|---|---|---|
| Christina Sondermann | 1 | 85 | 100 | 90 | 90 |
| Markus Huff | 2 | 5 | 0 | 5 | 0 |
| Martin Merkt | 3 | 10 | 0 | 5 | 10 |

Learning and Instruction 91 (2024) 101878

Contents lists available at ScienceDirect

Learning and Instruction

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

# Distracted by a talking head? An eye tracking study on the effects of instructor presence in learning videos with animated graphic slides

Christina Sondermann [a,*], Markus Huff [b,c], Martin Merkt [a]

[a] German Institute for Adult Education – Leibniz Centre for Lifelong Learning, Heinemannstraße 12-14, 53175, Bonn, Germany
[b] Department of Psychology, University of Tübingen, Schleichstraße 4, 72076, Tübingen, Germany
[c] Knowledge Media Research Center, Schleichstraße 6, 72076, Tübingen, Germany

ARTICLE INFO

*Keywords:*
Instructor presence
Talking head
Eye tracking
Video-based learning
Instructional design

ABSTRACT

*Background:* There are still unanswered questions concerning the optimal design of educational videos, for instance with regard to the visibility of a talking head next to the learning content. Although visible talking heads in educational videos may stimulate deeper processing, they can also distract from the visual learning content.
*Aims:* The current eye tracking study investigated how a talking head presented next to stepwise appearing (i.e., animated) graphic-based content affected learners' eye movements, learning, and ratings (e.g., social presence).
*Sample:* Ninety-six university students participated in the experiment.
*Methods:* We used a within-subjects design to vary whether a talking head was present in short learning videos with narrated slides.
*Results:* Our results revealed that the talking head both reduced fixation duration of the overall content and fixation duration and number of fixations of newly appearing content. Further, time to first fixation of newly appearing content was delayed by the talking head. However, the talking head did not affect learning outcomes (knowledge test and picture recognition task performance) and there was no mediation by the fixation duration of the content. Additionally, learners indicated higher social presence ratings for videos including a talking head, but we observed no differences in ratings of perceived learning, satisfaction, perceived difficulty, and invested mental effort due to the talking head.
*Conclusion:* The talking head strongly influenced learners' eye movements by reducing fixations of the learning content but did not cause any learning losses. We discuss implications and limitations of the findings also in view of future research.

## 1. Introduction

Learning with videos is becoming increasingly important - most recently enhanced by the COVID-19 pandemic. Video-based learning plays an important role in informal learning contexts, but teachers also increasingly implement learning videos in formal settings. Concerning the optimal design of such videos, considerable research and evidence are already available (e.g., Mayer, 2014a). However, regarding the question of whether a teaching person should be visible in learning videos next to the learning content, the findings are still inconsistent. Especially, there are still unresolved issues concerning the format *talking head*, which is widely used, for instance in lecture videos (e.g., Chen & Wu, 2015; Kizilcec, Bailenson, & Gomez, 2015). In this format, unlike other instructor presence formats, learners only see a video of a human instructor's head speaking into the camera, often in a corner of the learning video next to the learning content (e.g., video of a slideshow). Our study is the first to investigate the effects of a talking head in videos next to graphic-based content that appears gradually (animated slides) using eye tracking.

On the one hand, the talking head serves as a social cue and may encourage deeper processing of the learning content (cf. Mayer, 2014b) and increase the popularity of learning videos (e.g., Polat, 2022). But on the other hand, human faces can attract a considerable amount of visual attention (e.g., Kizilcec, Papadopoulos, & Sritanyaratana, 2014), raising the question of whether the visible instructor may distract from the visual learning content and thus interfere with learning (i.e.,

* Corresponding author.
*E-mail addresses:* sondermann@die-bonn.de (C. Sondermann), m.huff@iwm-tuebingen.de, markus.huff@uni-tuebingen.de (M. Huff), merkt@die-bonn.de (M. Merkt).

https://doi.org/10.1016/j.learninstruc.2024.101878
Received 30 March 2023; Received in revised form 10 November 2023; Accepted 12 January 2024
Available online 28 January 2024

split-attention effect; Ayres & Sweller, 2014). This potential visual distraction might be particularly challenging if the learning content next to the talking head contains a lot of visuals and the learners are not able to get all the relevant information from the audio track, creating a visual competition between the talking head and the learning content. In this study, we aim to systematically investigate which effect outweighs the other and gain a deeper insight into the attentional processes induced by talking heads in learning videos.

### 1.1. Background

Considering the theoretical background, social agency theory (SAT, Mayer, 2014b) emphasises the importance of social processes for learning. It assumes that social cues (e.g., a visible instructor) in the learning material can trigger a feeling of social presence, a social conversation schema, and social responses, which may in turn lead to deeper processing and enhanced learning. However, according to the embodiment principle (Mayer, 2021) of the cognitive theory of multimedia learning (CTML, Mayer, 2014a), instructors differ in their level of embodiment (high vs. low embodiment). This principle states that learners are more likely to learn deeply from instructors with a high level of embodiment (e.g., instructors showing gestures or body movements), than from instructors with a rather low embodiment (e.g., talking heads), implying that higher embodiment offers more potential benefits than lower embodiment. One reason for this discrepancy might be that decreasing embodiment results in the disappearance of referential cues such as gestures, shown to have learning-enhancing effects (e.g., Beege et al., 2020; Dargue, Sweller, & Jones, 2019).

Thus, the advantage of including a visible instructor may be particularly outweighed by potential disadvantages if only an instructor's face (i.e., a talking head) is included in the learning materials. In this respect, cognitive load theory (CLT, Sweller, 1999; Sweller, Ayres, & Kalyuga, 2011) and CTML (Mayer, 2014a) assume that learners' cognitive capacities are limited. In this context, CLT distinguishes between three types of cognitive load, whereby intrinsic cognitive load (ICL) results from the difficulty of the task and learners' prior knowledge (Sweller, van Merrienboer, & Paas, 1998). In contrast, the (unfavourable) design of the learning materials determines the extraneous cognitive load (ECL), which should be kept as low as possible. If learners manage ICL and ECL efficiently, they may use the remaining resources for elaborative cognitive processing, which is associated with learning and referred to as germane cognitive load (GCL). Accordingly, the coherence principle (Mayer & Fiorella, 2014) of the CTML would suggest that a visible instructor can result in an increase in extraneous cognitive load. Following this principle, all elements of the learning material that do not belong to the task itself should be excluded to reduce extraneous cognitive load.

Besides, in the context of instructor presence research, the split-attention principle (Ayres & Sweller, 2014) is often mentioned (e.g., Alemdag, 2022; Zhang, Xu, Pi, & Yang, 2022), since it suggests avoiding materials in which learners have to split their attention between multiple sources of information to prevent an increase in the extraneous load. However, it should be noted that the split-attention principle refers to different sources of information, "where each source of information is essential for understanding the material" (Ayres & Sweller, 2014, p. 206). Therefore, if the talking head provides no essential visual information necessary for learning, it is questionable whether the split-attention principle in its original form applies in this case.

Further, since research shows a fundamental preference for human faces (Gullberg & Holmqvist, 2006; Johnson, Dziurawiec, Ellis, & Morton, 1991), a visible talking head may distract from the visual learning content of a video. In cases where this visual content is particularly relevant to learning - because the relevant content is not (or cannot be) presented in the audio track - a talking head may result in learners missing important information.

### 1.2. Effects of instructor presence on visual attention

Numerous eye tracking studies investigating the effects of human instructors in educational videos found that an instructor attracts a considerable amount of visual attention (Colliot & Jamet, 2018; Kizilcec et al., 2014; Pi & Hong, 2016; van Wermeskerken, Ravensbergen, & van Gog, 2018; Wang & Antonenko, 2017; Wang, Antonenko, & Dawson, 2020; Zhang & Yang, 2022). This applies to both instructors with rather high embodiment (e.g., van Wermeskerken et al., 2018; Wang, Antonenko, & Dawson, 2020) and to those with a lower level of embodiment (e.g., talking heads, Kizilcec et al., 2014; Zhang & Yang, 2022). Further, the findings of Kizilcec et al. (2014) revealed that individuals differed in the percentage of time they spent looking at a talking head, ranging from 9% to 60%. Additionally, studies examining the effects of gaze guidance on visual attention suggested that an instructor looking directly into the camera - as is often the case in the talking head format - resulted in even greater attention on the face (e.g., Pi, Xu, Liu, & Yang, 2020; Polat, 2022; Stull, Fiorella, & Mayer, 2021; Wang, Pi, & Hu, 2019).

In line with the findings outlined above, several studies demonstrated that a visible instructor reduced visual attention on the learning content (Colliot & Jamet, 2018; van Wermeskerken et al., 2018; Wang, Antonenko, & Dawson, 2020; Zhang & Yang, 2022). Thus, learners spent less time looking at the learning content when an instructor was visible next to the content, both on the right and on the left (Zhang et al., 2022), with some of the corresponding studies showing very large effect sizes (e.g., $d = 3.06$, van Wermeskerken et al., 2018). For the talking head format, reduced attention to the learning content has previously only been observed for text-based content (Zhang & Yang, 2022), but to the best of our knowledge not for graphic-based content. Nevertheless, a detailed investigation of graphic-based learning content next to a talking head is of particular interest since in such materials not all information may also appear in the audio track and thus reduced visual attention might have a negative effect on learning.

However, for successful learning with educational videos, it may not only be crucial whether learners look at the learning content at all, but also when they look at specific content. Thus, for a successful integration of visual content and verbal explanation, it might be important that learners look at content when the narration refers to it (Mayer, 2014a; van Gog, 2014) because learners try to integrate verbal and pictorial information while watching videos (Schüler & Merkt, 2021). Therefore, van Wermeskerken et al. (2018) investigated whether a visible instructor affected how long, how often, and how fast learners looked at visual content that was currently discussed in the narration. They found that when an instructor was visible, learners fixated on the referenced content for a shorter time and less often, but not more slowly. As movements attract attention (Abrams & Christ, 2003), many designers of educational videos fade in their content step by step as soon as the narration refers to it to direct the learners' attention to the content that is currently important (cf. cueing principle, van Gog, 2014). However, to the best of our knowledge, it has not yet been investigated whether a talking head affects visual attention to this newly appearing content when it appears gradually.

### 1.3. Effects of instructor presence on learning outcomes

Although visible instructors in learning videos attract a lot of visual attention, the aforementioned studies did not observe any learning-impeding effects of instructors. While some of these studies reported no effects of the instructor on various learning measures (Kizilcec et al., 2014; van Wermeskerken et al., 2018), some even revealed learning-enhancing effects (Pi & Hong, 2016; Zhang & Yang, 2022), at least for certain learning materials or measures (e.g., retention of spoken explanations, Colliot & Jamet, 2018; recall performance for easy videos, Wang & Antonenko, 2017; transfer performance for difficult videos, Wang, Antonenko, & Dawson, 2020). Similarly, previous behavioural studies (i.e., without eye tracking) found mostly no effects (Homer,

Plass, & Blake, 2008; Ng & Przybylek, 2021; Sondermann & Merkt, 2023b) or beneficial effects on learning (Wang, Antonenko, Keil, & Dawson, 2020). This pattern is also reflected by a recent meta-analysis (Alemdag, 2022) and two recent reviews (Henderson & Schroeder, 2021; Polat, 2022) suggesting mostly positive and absent instructor effects on learning outcomes. Possible reasons for inconsistent findings include the heterogeneity of the learning materials and measures, but also the instructors' heterogeneity (for a systematic investigation of potential moderators, see Alemdag, 2022).

Whereas previous studies reported mainly positive or absent effects of instructor presence on learning measures, we recently observed a small but negative effect ($\eta_p^2 = 0.039$) of a talking head on recall of factual knowledge in an online study (Sondermann & Merkt, 2023a) using different slide types (graphic- and text-based slides) and presentation formats (animated and not animated slides). However, due to the online format, no eye movements were recorded in this study. Hence, it remained unclear whether the learning losses resulted from the learners paying less attention to the learning content.

In this regard, two studies using eye tracking (van Wermeskerken et al., 2018; Wang, Antonenko, & Dawson, 2020) already investigated whether visual attention spent on the visual learning content was related to learning outcomes. Whereas these previous studies examining instructors with rather high embodiment did not reveal any associations, a corresponding study for instructors with a lower level of embodiment (e.g., talking heads) does not exist, to the best of our knowledge. However, investigating instructors with low embodiment is of particular interest, since some potential advantages of high embodiment (e.g., referential cues such as gestures) should disappear while the disadvantages (e.g., distraction) should persist, leading to the risk of learning losses in this combination. Additionally, the question arises to what extent a talking head with low embodiment can act as a social cue by activating social schemata and influencing learning to the extent one would expect from an instructor with high embodiment.

A study that investigated eye movements during learning with an interactive tutoring system containing a virtual agent (Lallé, Taub, Mudrick, Conati, & Azevedo, 2017) already revealed associations between fixations on the agent and learning performance. However, learning performance differed depending on the students' achievement goals, with a high fixation rate on the agent resulting in higher learning for performance-oriented students, while a high fixation rate on the agent hindered learning for mastery-oriented students. Nevertheless, it is unclear in how far these results transfer to learning situations in which learners do not interact with the actual learning materials.

Further, since the instructor reduced attention to the learning content without interfering with learning in previous research (e.g., Kizilcec et al., 2014), the question arises to what extent attention to the visual learning content was necessary for successful completion of the respective knowledge tests. Since empirical evidence shows that visual attention affects recall of detailed information (Glaser & Schwan, 2015), it is conceivable that a talking head affects not only visual attention, but also the recognition of visual details. However, we are not yet aware of any study investigating whether a talking head affects learners' ability to remember visual details from videos.

### 1.4. Effects of instructor presence on subjective ratings

Besides the instructor effects on visual attention and learning measures, learners' subjective ratings are also a crucial factor for practice, especially in informal learning contexts, as learners may decide not to continue watching a video they do not like. Even though some studies did not observe effects of instructor presence on subjective ratings (Colliot & Jamet, 2018; Homer et al., 2008; Ng & Przybylek, 2021; Sondermann & Merkt, 2023b), there is ample evidence that a visible instructor positively affects satisfaction (Henderson & Schroeder, 2021; Sondermann & Merkt, 2023a; Wang & Antonenko, 2017; Wang, Antonenko, & Dawson, 2020), perceived learning (Kizilcec, Bailenson, & Gomez, 2015; Sondermann & Merkt, 2023a; Wang, Antonenko, & Dawson, 2020; Wilson et al., 2018), and perceived social presence (Sondermann & Merkt, 2023a). Correspondingly, both the recent meta-analysis and one recent review concluded that a visible instructor could increase motivation (Alemdag, 2022) and positive feelings (Polat, 2022). Further, Wang, Antonenko, and Dawson (2020) explored the relationships between visual attention measures and various video ratings and observed that learners reported higher levels of satisfaction the more they looked at the instructor.

Since, considering the CTML (e.g., coherence principle, Mayer & Fiorella; 2014), a visible instructor may result in increased extraneous cognitive load, various studies also examined the effects of instructor presence on different cognitive load measures. The findings in this regard are also mixed (Henderson & Schroeder, 2021). While some load-enhancing effects of instructor presence emerged (Homer et al., 2008; Hong, Pi, & Yang, 2018), other studies also revealed no (Colliot & Jamet, 2018; Ng & Przybylek, 2021) or even load-reducing effects of instructor presence (Kizilcec, Bailenson, & Gomez, 2015; Sondermann & Merkt, 2023a; Wang, Antonenko, Keil, & Dawson, 2020). For example, in a previous online study, we found that the talking head reduced ratings of perceived difficulty while increasing ratings of GCL (Sondermann & Merkt, 2023a). In line with CTML and CLT, the recent meta-analysis suggested a load-enhancing effect of the instructor (Alemdag, 2022).

### 1.5. The present research

Considering the unresolved questions outlined above, we aim to investigate how talking heads affect visual attention, learning outcomes, and learners' subjective ratings. Therefore, we systematically vary the presence of a talking head in short learning videos with narrated and animated slides. We focus on graphic-based contents on the slides because dynamic talking heads that do not show any gestures or body movements have so far only been shown to distract from text-based content (Zhang & Yang, 2022). Hence, we intend to examine whether the talking head also results in learners paying less attention to graphic-based content. In addition, following van Wermeskerken et al. (2018), the study aims to shed light on whether a talking head influences how long, how often, and how quickly learners fixate on new content that appears as soon as the narration refers to it.

In this study, we use similar learning materials and the same knowledge test as in a previous online study (Sondermann & Merkt, 2023a) to examine whether we can replicate the learning losses caused by a talking head in a laboratory setting and explain them with gaze data. Specifically, we explore whether potential learning losses due to a low embodied talking head can be explained by reduced attention to the learning content. Whereas previous research did not observe such correlations for instructors with higher levels of embodiment (van Wermeskerken et al., 2018; Wang, Antonenko, & Dawson, 2020), a positive correlation of attention to the learning materials and learning outcomes may be observed under conditions of low embodiment, because a present instructor without didactic functions (e.g., gestures, pointing) may not be able to compensate for potentially harmful effects.

Further, since many prior studies do not specify to what extent visual attention to the learning content was necessary for successful learning (i.e., whether items in knowledge tests could be answered based on the narration of the video), we included a picture recognition task (cf. Glaser & Schwan, 2015) as an additional learning measure in our study. In this task, learners had to recognise visual details from the videos without being able to extract the necessary information from the audio track, which is applicable due to the entirely graphic-based videos in this study. Finally, we also examined learners' subjective ratings (i.e., perceived learning, satisfaction, social presence, perceived difficulty, and invested mental effort) to further extend the existing heterogeneous body of evidence.

#### 1.5.1. Hypotheses

We preregistered all hypotheses and research questions on the open science framework (OSF, https://osf.io/rcde6/?view_only=fb6aa775e1b14d30826d2af29071bc51). Regarding visual attention, based on prior research (e.g., Colliot & Jamet, 2018; Kizilcec et al., 2014), we expected the overall learning content to be fixated longer (overall fixation duration of content) for videos without talking head than for videos including a talking head (Hypothesis 1). In addition, we preregistered to explore whether the talking head affects how long (fixation duration), how often (number of fixations), and how quickly (time to first fixation) learners fixate only the newly appearing content (Research Question 1, see also van Wermeskerken et al., 2018).

Concerning knowledge test performance, we expected to replicate the results from a previous online study (Sondermann & Merkt, 2023a) which found better learning outcomes for videos without talking head than for videos with talking head (Hypothesis 2). Analogously, we expected the talking head to also reduce the recognition of visual details from the videos (picture recognition task, Hypothesis 3), since visible instructors have been found to decrease viewing time of learning content in previous studies (e.g., Colliot & Jamet, 2018; van Wermeskerken et al., 2018). Additionally, we preregistered to explore whether the talking head effect on learning outcomes, i.e., knowledge test performance (Research Question 2) and picture recognition task performance (Research Question 3), depends on the amount of time the learning content is fixated.[1]

Regarding learners' subjective ratings, considering previous research (e.g., Henderson & Schroeder, 2021; Sondermann & Merkt, 2023a), we expected higher perceived learning (Hypothesis 4.1), satisfaction (Hypothesis 4.2), and social presence (Hypothesis 4.3) ratings for videos with talking head than for videos without talking head. In terms of cognitive load ratings, based on the findings of a previous study using similar materials (Sondermann & Merkt, 2023a), we expected lower ratings of perceived difficulty (Hypothesis 4.4) and higher ratings of invested mental effort (Hypothesis 4.5) for videos including a talking head.

## 2. Method

### 2.1. Participants and design

We recruited students of the local university via the student mailing list of the university and collected data of 101 participants. Based on preregistered exclusion criteria, we excluded $n = 2$ participants due to technical problems, $n = 2$ due to incomplete gaze samples (i.e., less than 80% complete gaze samples during the experimental procedure, Schüler, 2017, 2019), and $n = 1$ due to disturbance by a mobile phone during eye tracking. This resulted in a final sample size of $N = 96$ participants (73 female, $M_{age} = 22.61$ years, $SD_{age} = 3.78$). We had preregistered to stop data collection after two weeks, if we had collected data from at least 56 participants by then but aiming to include as many participants as possible in this period. The calculated minimum sample size was based on an a priori power analysis ($\alpha = 0.05$, 1-$\beta = 0.95$, $f = 0.2$) with G*Power 3.1.9.4 (Faul, Erdfelder, Lang, & Buchner, 2007) to detect a talking head effect from the previously conducted online study (Sondermann & Merkt, 2023a). As preregistered, we stopped data collection on May 13, 2022.

Participants watched eight short videos on eight different topics. We used a within-subjects design to vary whether a talking head was present as the independent variable, with each participant watching four different videos with and four different videos without a talking head. Further, we balanced between participants which video topic was presented with and without talking head and whether the talking head was shown on the right or left to avoid confounds of the presence of talking heads with contents or position. Since we further balanced two different video orders, this resulted in a total of eight balancing sequences.

### 2.2. Materials

The learning materials consisted of eight short educational videos with narrated slides that provided factual knowledge on two geographical (Burkina Faso, South Georgia and the South Sandwich Islands), two physical (Radiocarbon dating, Foucault pendulum), two biological (Ficus lyrata, Blackcap), and two historical topics (Wilhelmine Reichard, Rudolf Höß). All videos lasted about 1 min, and each consisted of one slide with graphic content (e.g., pictures, maps) only. The slides were animated with the content gradually appearing on the slide as soon as the soundtrack referred to it. The appearance of the content did not follow a uniform structure (e.g., from top to bottom). For instance, for the topic Burkina Faso, the content showed up in different places on the map (first all neighbouring countries, then the three largest cities), whereas for the topic Blackcap, a larger image of the bird was first displayed in the middle before additional images appeared in the top right and top left.

To balance between participants which topic is presented with and without talking head, we produced versions with and without a female talking head for each video topic. In versions with talking head, the narration was provided by the talking head. In versions without talking head, the soundtrack was completely identical. In addition, we balanced whether the talking head appeared on the bottom right or left corner of the video. For videos with talking head, we produced one version in which the talking head appeared on the right and the content rather on the left, and one in which the talking head appeared on the left and the content rather on the right. Correspondingly, for the videos without a talking head, there was one version in which the content appeared rather on the right and one in which the content appeared rather on the left (see Fig. 1). We also systematically balanced this content position in conditions without talking head to rule out the possibility that potentially occurring effects resulted from a slightly different position of the content. Completely crossing these elements (talking head vs. no talking head; right vs. left) resulted in four different versions for each of the eight video topics, meaning that we produced a total of 32 different videos.

### 2.3. Apparatus

Videos were presented via Tobii Pro Lab (Version 1.123) on a monitor with a resolution of 1920 x 1080. We recorded participants' eye movements with screen-based Tobii Pro Nano eye trackers (60 Hz). To improve data quality and to maintain a constant distance from the screen (65 cm), we used chinrests on which participants placed their heads during video playback. The calibration procedure consisted of a nine-point calibration and a four-point validation and was repeated when the average deviation for accuracy or precision were higher than 1.0° ($M_{\text{validation accuracy}} = 0.35°$, $SD_{\text{validation accuracy}} = 0.20$, $M_{\text{validation precision}} = 0.26°$, $SD_{\text{validation precision}} = 0.17$). The applied Tobii I-VT fixation gaze filter used a maximum velocity of 30°/s and a minimum fixation duration of 60 ms to determine a fixation.

### 2.4. Measures

#### 2.4.1. Individual prerequisites

We asked for participants' self-reported prior knowledge about the topics covered in the videos on 7-point Likert scales ranging from 1 (*no prior knowledge*) to 7 (*a lot of prior knowledge*) with one question per topic (e.g., "How much do you know about the African country Burkina Faso?"). In addition, participants indicated their prior interest in the

[1] Initially, we preregistered moderation analyses as exploratory analyses for these research questions but decided to use mediation analyses after receiving collegial feedback.

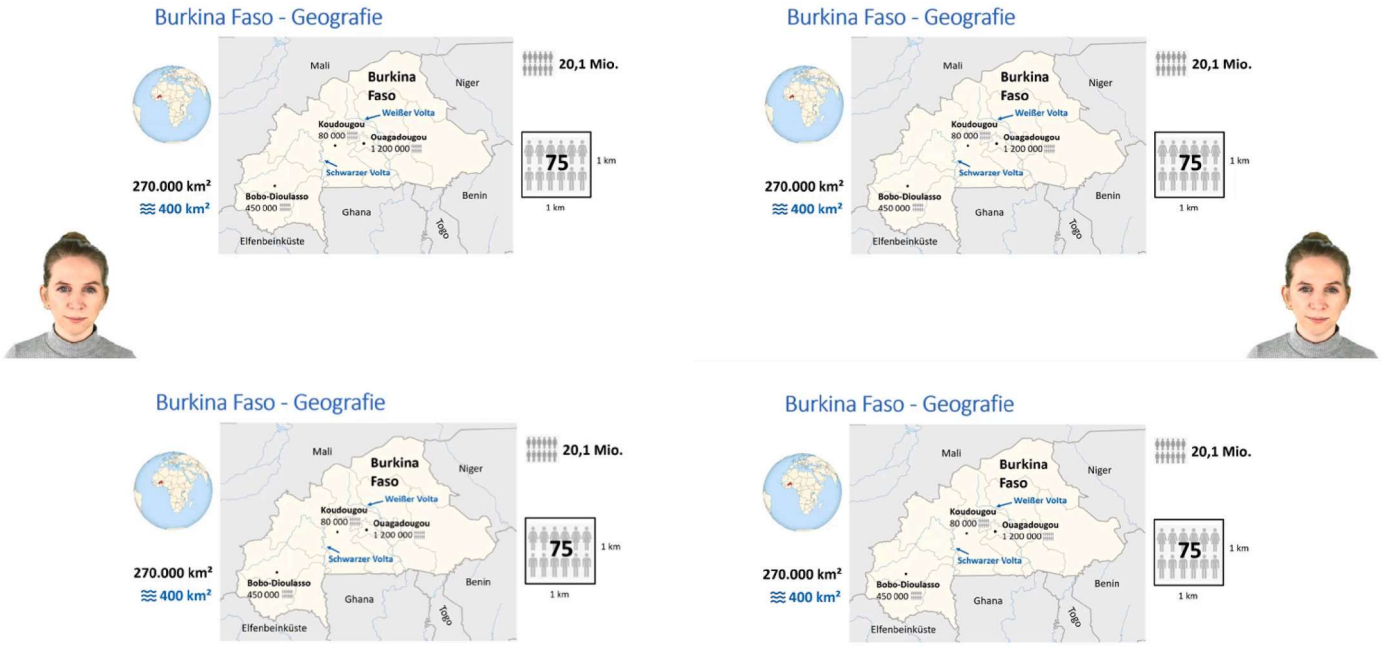



**Fig. 1.** Screenshots or videos with (top) and without talking head (bottom). Depending on the balancing, the talking head was displayed on the left (and the content more on the right) or on the right (and the content more on the left).

eight topics (e.g., "How interested are you in the African country Burkina Faso?") on a 7-point Likert scale from 1 (*not at all*) to 7 (*very interested*).

### 2.4.2. Visual attention

*2.4.2.1. Overall fixation duration of content.* We defined areas of interest (AOIs) for the learning contents in the videos. As the content appeared on the slides gradually, the area for the overall fixation duration of content increased over time. Hence, as soon as a new content appeared, the overall area was extended by a corresponding AOI. Fixation durations of content for videos with and without talking head were calculated for each person. Since the different videos were of different lengths, we divided the overall fixation duration by the duration of the respective videos for all analyses.

*2.4.2.2. Fixation duration of newly appearing content.* Further, we examined how long learners spent fixating only the new content that appeared on the slides. The respective AOIs were identical to those AOIs used to determine the overall fixation duration of content but were only active from the visual onset of each content to the onset of the subsequent content (time limited AOIs). Again, we calculated content fixation durations for videos with and without talking head and divided them by the duration of the respective videos.

*2.4.2.3. Number of fixations of newly appearing content.* We also recorded the number of fixations of newly appearing content, i.e., how often learners looked at the new content using the identical, time limited AOIs as for the fixation duration of newly appearing content. We calculated numbers of fixations for videos with and without talking head and divided them by the duration of the respective videos. Since the eye tracking data was provided in milliseconds, we multiplied the resulting values by 1000 to obtain the numbers of fixations per second (instead of per millisecond).

*2.4.2.4. Time to first fixation of newly appearing content.* In addition, we collected the time to first fixation of the newly appearing content, i.e., how long learners took on average to look at a piece of content after it appeared. We calculated the average time to first fixation for each video by adding up the time to first fixations for the individual AOIs and dividing them by the respective number of AOIs. Since the respective AOIs were only active for a limited time, not all AOIs were fixated during their active phase or fixated at all. Therefore, in these two cases, the maximum duration of the respective AOI (from visual onset of the content to the onset of the subsequent content) was used.[2] For each person, an average time to first fixation was calculated for videos with and without talking head. Again, this was based on the same time limited AOIs as the two previous measures.

### 2.4.3. Learning outcomes

*2.4.3.1. Knowledge test.* In our self-designed knowledge test, participants answered three open recall questions for each of the eight topics, resulting in a total of 24 questions. Correct answers consisted of single (or few) words or numbers addressing facts from the videos (e.g., "Name three neighbouring countries of Burkina Faso."). Two independent raters assessed their correctness by using a coding scheme, assigning 1 point for correct answers, 0.5 for partially correct answers, and 0 points for incorrect answers. Inter-rater agreement was almost perfect for all 24 items, ranging from Cohen's $\kappa = 0.89$ to $\kappa = 1$. Discrepancies were resolved by the first author. Points were summed up per condition.

*2.4.3.2. Picture recognition task.* In the picture recognition task, learners had to recognise detailed content from the videos that had only been presented visually. The task consisted of three multiple choice questions per topic (24 items in total). Learners had to choose the picture that appeared in the video from four options, with one correct answer and

[2] Instead of replacing values, we also applied alternative measures: We (1) included only those instances in which an AoI was fixated at all, (2) or was fixated in time, and (3) considered in which time 50% of the participants had fixated a respective AoI per condition (T50, see Hooge & Camps, 2013; Meier et al., 2023; van Wermeskerken et al., 2018). Since all measures led to the same conclusions, we refrain from presenting them separately in the manuscript, but refer to Appendix A, where the corresponding results for all measures can be found.

three distractors for each item (see Fig. 2 for an example). They received one point for each correct answer and the points were summed up to a total score (each for videos with and without talking head). Due to technical issues, the picture recognition task responses on the blackcap topic were not recorded. Therefore, the value for this topic was replaced by all participants' overall average recognition task performance based on the other seven topics ($M = 2.34$). Since different learners saw the blackcap topic equally often in different video versions due to balancing, this procedure has no systematic influence in favour of one particular video format.

#### 2.4.4. Subjective ratings

*2.4.4.1. Perceived learning and satisfaction.* Learners indicated how much they learned from a video on a 7-point Likert scale ranging from 1 (*nothing at all*) to 7 (*a lot*) adapted from Kizilcec et al. (2015) and Wang and Antonenko (2017). Further, they assessed their satisfaction with the learning materials on a 7-point Likert scale from 1 (*very dissatisfied*) to 7 (*very satisfied*) adapted from Wang and Antonenko (2017).

*2.4.4.2. Social presence.* Social presence was captured with three items on a 7-point Likert scale from 1 (*strongly disagree*) to 7 (*strongly agree*). We adapted two items (e.g., "I felt that the instructor was present") from a questionnaire used by Kizilcec et al. (2015). A third item ("I felt like the instructor addressed me personally") was taken from the PSI process scales (Schramm & Hartmann, 2008) already applied by Beege, Schneider, Nebel, and Rey (2017). Ratings for the three items were averaged to an overall score for videos with talking head (Cronbach's $\alpha = 0.93$) and an overall score for videos without talking head (Cronbach's $\alpha = 0.94$).

*2.4.4.3. Invested mental effort and perceived difficulty.* After each video, learners indicated how much mental effort they invested to follow the video (adapted from Paas, 1992) on a 7-point Likert scale ranging from 1 (*not at all*) to 7 (*a great deal*). We used this item in its active form, since it can be assumed that its measurement is more likely to capture the GCL facet (Klepsch & Seufert, 2021). Further, learners assessed how easy or difficult it was to understand the video on a 7-point Likert scale from 1 (*extremely easy*) to 7 (*extremely difficult*) adapted from Kalyuga, Chandler, and Sweller (1999). We applied this measure because it has also been used previously as a measure of mental load (Kalyuga et al., 1999) and was strongly associated with the three-item measure of ECL by Klepsch, Schmitz, and Seufert (2017) in a previous study (Sondermann & Merkt, 2023a).

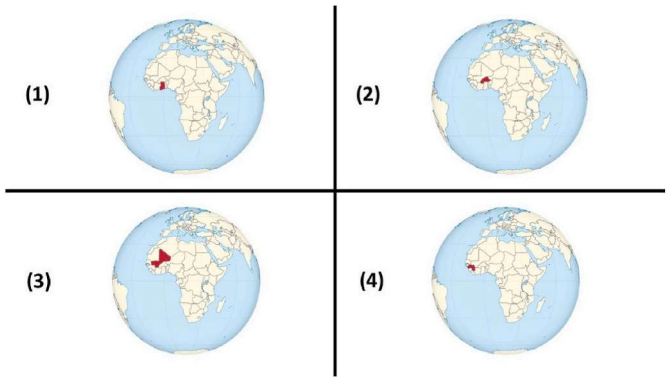



**Fig. 2.** An example item from the picture recognition task on the topic Burkina Faso. Participants were asked which of the four pictures they had seen in the respective video and had to choose one of the four pictures. Picture credits: This graphic contains elements which have been taken from the following graphics (CC BY-SA 3.0): (1) https://de.wikipedia.org/wiki/Datei:Ghana_on_the_globe_(Africa_centered).svg, (2) https://de.wikipedia.org/wiki/Datei:Burkina_Faso_on_the_globe_(Africa_centered).svg, (3) https://de.wikipedia.org/wiki/Datei:Mali_on_the_globe_(Africa_centered).svg, (4) https://de.wikipedia.org/wiki/Datei:Guinea_on_the_globe_(Africa_centered).svg.

### 2.5. Procedure

The experiment was conducted at a German university. Up to three participants took part at the same time in the laboratory, separated by partition walls. After obtaining participants' informed consent, we randomly assigned them to one of eight balancing sequences. First, participants assessed their prior knowledge and interest about the eight video contents using MediaLab (Version 2016.1.108). Afterwards, they switched to another laptop which was used for the eye tracking part (see Section 2.3). Once the participants were seated correctly and had placed their head on a chinrest, the experimenter started the calibration procedure. Subsequently, all participants watched eight learning videos which started automatically and did not enable any user interactions (e.g., pause, replay, pace). After each video, we collected ratings of invested mental effort, perceived difficulty, perceived learning, satisfaction, and social presence. After completing the eye tracking part and returning to the laptop with MediaLab, the knowledge test and the picture recognition task followed. Finally, participants answered some demographic and control questions (e.g., familiarity of the instructor) and were debriefed. For their participation, participants received either course credit or financial compensation (10€). The procedure was approved by the local ethics committee.

### 2.6. Analytical strategy

The study used a complete within-subjects design in which each participant was confronted with each of the different experimental conditions, thus ruling out that individual prerequisites and differences may confound potential effects of conditions. Further, participants were independent from each other. Therefore, we used repeated-measure ANOVAs with the independent variable talking head as a standard procedure to analyse within-subjects designs. In addition, we conducted mediation analyses to examine whether the talking head's effect on learning outcomes was mediated by the fixation duration of content. Due to our repeated measures design, we used the MEMORE_2.1 macro in SPSS (Montoya & Hayes, 2017) which enables mediation analyses with this data structure. Data and scripts can be downloaded at OSF (https://osf.io/4uwx6/?view_only=47a628aba0e54eafa7a0887b60af4266).

## 3. Results

Table 1 displays descriptive results for the visual attention measures. Means and standard deviations of the learning outcomes and all rating scales can be found in Table 2. Correlations between all dependent variables are shown in Table B.1 in Appendix B.

**Table 1**
Means (and standard deviations in parentheses) for the visual attention measures.

| Variable | Overall content | Newly appearing content | | |
|---|---|---|---|---|
| | Fixation duration[a] | Fixation duration[a] | Number of fixations (per second)[a] | Time to first fixation (in ms) |
| Talking head | 0.71 (0.10) | 0.45 (0.10) | 1.05 (0.22) | 1211.81 (347.99) |
| No talking head | 0.83 (0.09) | 0.54 (0.10) | 1.21 (0.23) | 1089.29 (237.84) |

*Note.* $N = 96$.
[a] Since the different videos were of different lengths, we report the fixation durations and number of fixations divided by the duration of the respective videos.

**Table 2**
Means (and standard deviations in parentheses) for learning outcomes and rating scales.

| Variable | Knowledge test | Picture recognition | Perceived learning | Satisfaction | Social presence | Invested mental effort | Perceived difficulty |
|---|---|---|---|---|---|---|---|
| Talking head | 4.94 (2.07) | 9.41 (1.68) | 4.35 (0.93) | 4.59 (1.01) | 3.39 (1.43) | 4.57 (1.51) | 2.40 (0.99) |
| No talking head | 5.07 (2.10) | 9.34 (1.63) | 4.29 (0.89) | 4.54 (0.98) | 2.55 (1.24) | 4.56 (1.43) | 2.46 (1.05) |

*Note.* $N = 96$.

### 3.1. Preliminary analyses on individual prerequisites

There were no differences in self-reported prior knowledge ($M = 1.31$, $SD = 0.40$) or interest ($M = 3.14$, $SD = 0.90$) between videos presented with and without a talking head, both $F < 1$.

### 3.2. Visual attention

In line with Hypothesis 1, overall fixation durations of content were shorter when a talking head was visible than when no talking head was visible, $F\,(1,95) = 281.59$, $p < .001$, $\eta_p^2 = 0.748$. Regarding the newly appearing content, fixation duration was also lower with talking head than without talking head, $F\,(1,95) = 112.54$, $p < .001$, $\eta_p^2 = 0.542$. Additionally, the number of fixations of this newly appearing content was lower when a talking head was present than when no talking head was present, $F\,(1,95) = 88.60$, $p < .001$, $\eta_p^2 = 0.483$. Finally, the average time to first fixation of newly appearing content was higher for videos with talking head than for videos without talking head, $F\,(1,95) = 45.62$, $p < .001$, $\eta_p^2 = 0.324$.

### 3.3. Learning outcomes

Contrary to Hypotheses 2 and 3, there were no talking head effects on either knowledge test performance or picture recognition task performance, both $F < 1$.

### 3.4. Subjective ratings

Contrary to Hypotheses 4.1, 4.2, 4.4, and 4.5, analyses revealed no talking head effects on perceived learning, satisfaction, perceived difficulty, and mental effort, all $F < 1$. However, in line with Hypothesis 4.3, we found a positive talking head effect on social presence, with higher ratings for videos with talking head than for those without talking head, $F\,(1,95) = 79.12$, $p < .001$, $\eta_p^2 = 0.454$.

### 3.5. Mediation analyses

Mediation analyses were conducted using the MEMORE_2.1 macro in SPSS (Montoya & Hayes, 2017) to investigate if the talking head effects on learning outcomes were mediated by fixation duration on the learning content. With respect to the dependent variables, we conducted mediation analyses for both knowledge test and picture recognition task performance. Regarding the mediator, we performed these analyses for both the fixation duration of newly appearing content and the overall fixation duration of content. This resulted in a total of four different mediation models (see Fig. 3 for an illustration). With respect to the overall fixation duration of content, analyses revealed no significant mediations, neither for knowledge test nor for picture recognition task performance, for both overall models $F < 1$. Similarly, fixation duration of newly appearing content did not mediate the talking head effect on knowledge test performance, $F\,(2,93) = 1.52$, $p = .224$, or picture recognition task performance, $F\,(2,93) = 1.00$, $p = .371$. Since none of these analyses provided an overall significant mediation model, we refrain from presenting the individual paths for reasons of space.

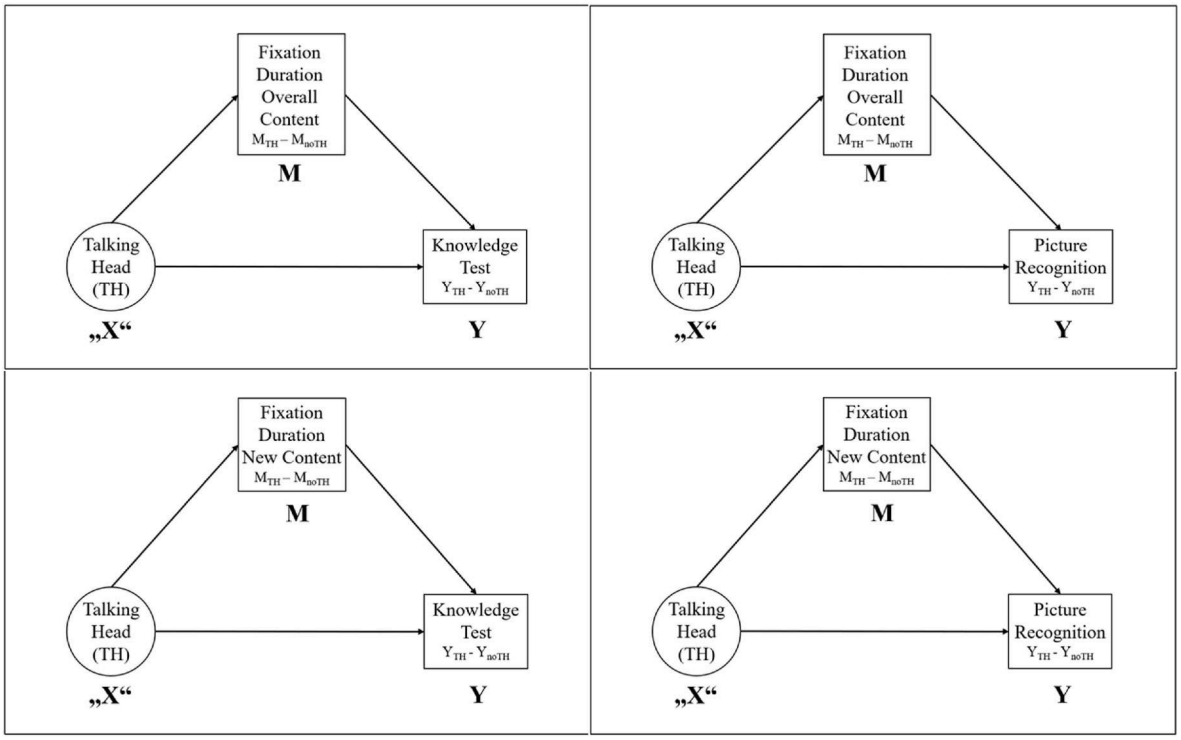



**Fig. 3.** An illustration of the mediation models investigated.

## 4. Discussion

In summary, our findings demonstrate that a visible talking head led learners to fixate on the overall learning content for a shorter time. They also fixated on the newly appearing content for a shorter period of time, less frequently, and more slowly when a talking head was visible next to the content. Nevertheless, the talking head did not affect learning outcomes, neither knowledge test nor picture recognition task performance, and the relationship was not explained by the fixation time of the content. Further, a visible talking head significantly increased ratings of social presence but had no effect on subjective ratings of perceived learning, satisfaction, difficulty, and invested mental effort.

Taken together, our findings reveal that talking heads have a strong influence on visual attention. In line with Hypothesis 1 and previous research (e.g., Colliot & Jamet, 2018), a visible talking head resulted in shorter viewing times of visual content. Our findings extend the previous evidence (e.g., van Wermeskerken et al., 2018) in that we observe reduced attention on graphic-based content presented next to a talking head.

In addition, we found that a talking head led to learners fixating the newly appearing content shorter, less often, and more slowly (Research Question 1). Hence, our results on fixation duration and numbers of fixation support the findings of van Wermeskerken et al. (2018) and extend them by demonstrating that a talking head can also delay the fixation of newly appearing content. The fact that we were able to observe such a delay caused by the instructor for the first time, unlike van Wermeskerken et al. (2018), may lie in the different design of the materials. Whereas the slide content in the videos of van Wermeskerken et al. (2018) appeared procedurally from top to bottom, the content in our study did not appear with such a rigid structure making it difficult to anticipate where the next content would show up. Accordingly, orientation on the slides was probably easier in the material of van Wermeskerken et al. (2018), allowing the learners to quickly find the current content even with an instructor present. In our materials, however, the area in which new content could potentially appear was larger, so that distractions caused by the talking head were possibly more apparent. This novel finding, which deviates from previous research, thus suggests that delayed fixation of newly appearing content by an instructor can occur especially with materials that are not clearly structured visually. This has important implications for the design of learning materials, which should be investigated in more detail in future research.

However, although the talking head strongly decreased the visual attention on the learning content, it had no impact on learning outcomes. Contrary to Hypothesis 2 and a previous online study (Sondermann & Merkt, 2023a), the talking head did not impair knowledge test performance (i.e., recall of factual information). Similarly, contrary to Hypothesis 3, the talking head had no effect on picture recognition task performance, although the task required visual attention to details of the learning content. Additionally, the exploratory mediation analyses addressing Research Questions 2 and 3 did not provide any evidence that a talking head effect on learning outcomes is explained by fixations of the learning content, which is consistent with previous research (van Wermeskerken et al., 2018; Wang, Antonenko, & Dawson, 2020).

The absence of a talking head effect on learning outcomes despite reduced visual attention to the learning content was also reported in other studies on instructor presence (Kizilcec et al., 2014; van Wermeskerken et al., 2018). One possible explanation for this pattern of results is that the involved processes (i.e., deeper processing through social cue, increased extraneous cognitive load due to extraneous processing) may neutralise each other, resulting in no overall talking head effects on learning. Hence, the learning-enhancing effects of the talking head as a social cue (deeper processing through agency, Mayer, 2014b), which also appeared in some other studies with instructor presence (Colliot & Jamet, 2018; Pi & Hong, 2016; Wang & Antonenko, 2017; Wang, Antonenko, & Dawson, 2020; Zhang & Yang, 2022), may have compensated for the learning losses postulated by CLT and CTML due to an increase of extraneous cognitive load (Ayres & Sweller, 2014; Mayer, 2014a). This reasoning is supported by the fact that the talking head strongly increased learners' ratings of social presence in line with SAT (Mayer, 2014b). However, the reasoning is contradicted by the fact that we observed no differences in cognitive load measures due to the talking head.

In view of the missing talking head effects on cognitive load measures (e.g., difficulty), it is worth noting that split-attention effects are not necessarily reflected in measures of extraneous cognitive load (Schroeder & Cenkci, 2020). Beyond this, with respect to the split-attention principle, the question arises whether this principle applies to our material combination at all. As already suggested in Section 1.1, the talking head provides no essential visual information necessary for learning, implying that learners do not need to integrate visual information of the talking head with the visual learning content presented on the slides. However, this would be a precondition for the occurrence of split-attention effects if we interpret this effect strictly according to its definition (Ayres & Sweller, 2014). Nevertheless, the present study provides empirical evidence that learners split their visual attention between the talking head and the learning materials, which can also lead to important details of the visual learning content being missed. Hence, it cannot be ruled out that this particular form of split-attention in which elements irrelevant to learning attract visual attention might be even more detrimental (e.g., with more visually complex learning content) than classic split-attention effects in which learners need to split their attention between relevant sources of information.

Considering the theoretical implications of our findings, we might face the question to what extent the mechanisms postulated by the SAT (Mayer, 2014b) can be regarded as automatic processes. For instance, the SAT suggests that higher social presence resulting from the inclusion of a social cue leads to higher elaboration. However, considering our results, this does not seem to be an automatic process because we did not observe any correlation between social presence and invested mental effort (see Appendix B). Nevertheless, in this respect, we would like to emphasise that the measure of invested mental effort we used to capture GCL consisted of only one item (see Section 4.1). However, if such associations remain absent also in studies with other measures of GCL, future research could take a closer look at the specific mechanisms resulting in deeper processing.

Particularly regarding the picture recognition task, the question arises why we observed no learning losses due to the talking head despite the evident distraction at the level of visual attention. At this point, however, it should be noted that across conditions, a positive correlation between number of fixations of newly appearing content and picture recognition task performance emerged, whereas no such correlation showed up between content fixation duration and picture recognition task performance (see Appendix B). Since the successful completion of the picture recognition task required looking at the learning content and we can therefore exclude that learners were able to extract the correct answers from the audio track, it is possible that only very brief fixations of the learning content were sufficient to successfully complete the picture recognition task. For example, previous research (Eitel, Scheiter, & Schüler, 2012) has shown that presentation times of 600 ms were sufficient to extract the gist of a picture. Although our picture recognition task asked for visual details, it cannot be ruled out that the participants were able to answer some items successfully despite being distracted by the talking head. Therefore, it might be interesting to investigate the talking head effects for graphical content with a higher visual complexity (see Meier et al., 2023).

In general, we can also note that we observed different effects in this laboratory study compared to a previous online study (Sondermann & Merkt, 2023a), not only in terms of knowledge test performance. Additionally, in contrast to Hypotheses 4.1, 4.2, 4.4, and 4.5 and the results of our online study (Sondermann & Merkt, 2023a), the talking head had no effect on learners' ratings of perceived learning, satisfaction, perceived difficulty, and invested mental effort. Only for social

presence we observed a strong increase in ratings for videos including a talking head, in line with Hypothesis 4.3 and previous research (Sondermann & Merkt, 2023a). These differences in the pattern of findings despite similar materials may be due to the different study settings. After all, watching a video in the laboratory while resting the head on a chinrest, knowing that one's eye movements are being recorded while an experimenter is in the room, represents a completely different setting than watching the same video at home in a familiar environment without supervision. Clearly, these contextual factors have implications for internal validity (higher in the laboratory) and ecological validity (higher in the online experiment).

However, beyond this methodological component, it is also conceivable that effects of video design differ depending on study setting. Hence, in future research, it would be useful to examine the effects of certain contextual factors. Supporting this argument, Alemdag's (2022) meta-analysis already revealed a moderating effect of study setting on instructors' effects on knowledge acquisition, reporting learning-enhancing effects of instructors in the laboratory. In contrast, for online studies, the meta-analysis suggests a - non-significant - but descriptively negative effect of the instructor on learning. It is conceivable that due to the higher degree of formality in the laboratory setting, learners invest more effort and thus compensate for potential detrimental effects of the instructor, whereas they are more likely to be distracted by an instructor in an online setting. In line with this reasoning, Zhang, Miller, Sun, and Cortina (2020) found that increasing allocations of fixations to the instructor's image were associated with mind-wandering during video lectures.

### 4.1. Limitations and directions for future research

When considering the findings of the present study, certain limitations should also be taken into account. For example, by asking for ratings of invested mental effort (adapted from Paas, 1992) and perceived difficulty (Kalyuga et al., 1999) with one item each, our assessment of different types of cognitive load may be considered suboptimal. Particularly, these measures cannot cover all three types of cognitive load. However, the perceived difficulty rating we used was strongly associated with the three-item measure of extraneous load by Klepsch et al. (2017) in a previous study (Sondermann & Merkt, 2023a) and resulted in the same talking head effects, suggesting that both measures might reflect similar aspects, even though we cannot distinguish between ICL and ECL with the measure of perceived difficulty. In addition, we applied the invested mental effort item in its active form to depict the GCL facet (Klepsch & Seufert, 2021). In this study, we decided against a more detailed survey of the different sources of cognitive load with several items (e.g., Klepsch et al., 2017; Krieglstein, Beege, Rey, Sanchez-Stockhammer, & Schneider, 2023) because we had to record the respective ratings after each of the eight videos due to our within-subjects design. To keep the eye tracking part as short as possible, we did not use overly time-consuming measures with multiple items since the main focus of our study was on eye movement measures and learning outcomes.

Additionally, it is important to note that the videos investigated were rather short, implying that we cannot generalise our findings to longer videos without further research. Although initial results (e.g., van Wermeskerken et al., 2018) already suggest that visual attention on the instructor does not decrease over time even with longer videos, future research could take a closer look at learners' attentional processes over time when watching longer videos, including potential adaptive mechanisms. For example, in a study investigating a gaze-reactive tutoring system (D'Mello, Olney, Williams, & Hays, 2012), an eye tracker monitored students' gaze patterns to identify at what point the students were zoning out. In these cases, the instructor prompted students to pay attention, which resulted in reorientation of attention and improvement in deeper learning. Similarly, it might be possible to fade out a talking head in adaptive learning videos when learners look at the face for a certain time. Such an adaptive fade-out would be especially beneficial against the background of the previously mentioned finding that increasing attention on the instructor is associated with mind-wandering (Zhang et al., 2020).

### 4.2. Conclusion

Overall, we can conclude that a talking head causes a strong distraction on the level of visual attention (e.g., content fixation duration), without this visual distraction being reflected in the learning differences. Since we could not identify any learning losses, considering this study there is little to argue against the inclusion of a talking head in learning videos. However, apart from the increase in social presence, no clear advantages of including a talking head were evident. In addition, the study setting may be considered as a potential boundary condition, implying that learning losses due to a talking head cannot be ruled out in more informal settings.

## CRediT authorship contribution statement

**Christina Sondermann:** Writing – original draft, Writing – review & editing, Conceptualization, Methodology, Project administration, Investigation, Formal analysis, Data curation. **Markus Huff:** Writing – review & editing, Conceptualization, Methodology, Project administration, Supervision. **Martin Merkt:** Writing – review & editing, Conceptualization, Methodology, Project administration, Supervision.

## Declarations of competing interest

None.

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

The study received an ethics approval from the local ethics committee (ethikkommission@die-bonn.de) at the German Institute for Adult Education - Leibniz Centre for Lifelong Learning (approval number: DIE-LEK 2019–012). Informed consent was obtained from all individual participants.

## Acknowledgements

We would like to thank our student research assistants Lara Kläffling and David Rohde for supporting our data collection and Inka Krüger and Claas Vey for coding the students' answers. We would also like to thank Julia Granderath for her effort as talking head.

## Appendix A

*Time to first fixation of newly appearing content*

Besides replacing the values with the maximum duration if an AoI was not fixated at all or was fixated too late, we applied three additional measures for the analyses on time to first fixation. First, we included in our analyses only those instances in which an AoI was fixated at all, while instances in which an AoI was not fixated were not included. According to the chosen measure reported in the Method section, this measure also resulted in a significant effect of the talking head, with newly appearing content being fixated more slowly when a talking head was visible ($M$ =

1143.37, $SD = 144.21$) than when no talking head was visible ($M = 1046.57$, $SD = 104.18$), $F(1,95) = 54.30$, $p < .001$, $\eta_p^2 = 0.364$.

Second, we included in our analyses only those instances in which an AoI was fixated in time, while instances in which an AoI was fixated too late or not at all were not included. Again, when using this measure, newly appearing content was fixated more slowly when there was a talking head next to it ($M = 1121.38$, $SD = 127.75$) than when there was no talking head next to it ($M = 1031.66$, $SD = 101.07$), $F(1,95) = 56.55$, $p < .001$, $\eta_p^2 = 0.373$.

Third, we applied the T50 measure (see Hooge & Camps, 2013; Meier et al., 2023; van Wermeskerken et al., 2018), that determines in which time 50% of the participants had fixated a respective AoI per condition. Corresponding to all other measures, the time to first fixation of newly appearing content was higher with talking head ($M = 887.78$, $SD = 57.10$) than without talking head ($M = 864.18$, $SD = 47.37$), $F(1,95) = 17.74$, $p < .001$, $\eta_p^2 = 0.157$.

## Appendix B

**Table B.1**
Correlations Between all Dependent Variables Included in the Study

| Variable | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 1. Fixation duration (overall content) | – | | | | | | | | | | |
| 2. Fixation duration (newly appearing content) | .861*** | – | | | | | | | | | |
| 3. Number of fixation (newly appearing content) | −.002 | .119 | – | | | | | | | | |
| 4. Time to first fixation (newly appearing content) | −.516*** | −.667*** | −.538*** | – | | | | | | | |
| 5. Knowledge test | .065 | .092 | .154 | −.178 | – | | | | | | |
| 6. Picture recognition | .038 | .043 | .236* | −.073 | .383*** | – | | | | | |
| 7. Perceived learning | .194 | .231* | −.076 | −.105 | .096 | .105 | – | | | | |
| 8. Satisfaction | .140 | .142 | −.124 | −.127 | .047 | .022 | .713*** | – | | | |
| 9. Social presence | .110 | .084 | −.013 | −.146 | −.047 | −.018 | .199 | .421*** | – | | |
| 10. Invested mental effort | .016 | −.051 | .112 | .046 | .037 | .068 | .161 | .084 | −.065 | – | |
| 11. Perveived difficulty | −.076 | −.093 | .104 | .030 | −.159 | −.093 | −.060 | −.036 | .160 | .186 | – |

*Note.* $N = 96$.
$^*p < .05$. $^{**}p < .01$. $^{***}p < .001$.

## References


Abrams, R. A., & Christ, S. E. (2003). Motion onset captures attention. *Psychological Science, 14*, 427–432. https://doi.org/10.1111/1467-9280.01458

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

Ayres, P., & Sweller, J. (2014). The split-attention principle in multimedia learning. In R. E. Mayer (Ed.), *The Cambridge handbook of multimedia learning* (2nd ed., pp. 206–226). Cambridge University Press.

Beege, M., Ninaus, M., Schneider, S., Nebel, S., Schlemmel, J., Weidenmüller, J., et al. (2020). Investigating the effects of beat and deictic gestures of a lecturer in educational videos. *Computers & Education, 156*, Article 103955. https://doi.org/10.1016/j.compedu.2020.103955

Beege, M., Schneider, S., Nebel, S., & Rey, G. D. (2017). Look into my eyes! Exploring the effect of addressing in educational videos. *Learning and Instruction, 49*, 113–120. https://doi.org/10.1016/j.learninstruc.2017.01.004

Chen, C.-M., & Wu, C.-H. (2015). Effects of different video lecture types on sustained attention, emotion, cognitive load, and learning performance. *Computers & Education, 80*, 108–121. https://doi.org/10.1016/j.compedu.2014.08.015

Colliot, T., & Jamet, É. (2018). Understanding the effects of a teacher video on learning from a multimedia document: An eye-tracking study. *Educational Technology Research & Development, 66*, 1415–1433. https://doi.org/10.1007/s11423-018-9594-x

D'Mello, S., Olney, A., Williams, C., & Hays, P. (2012). Gaze tutor: A gaze-reactive intelligent tutoring system. *International Journal of Human-Computer Studies, 70*, 377–398. https://doi.org/10.1016/j.ijhcs.2012.01.004

Dargue, N., Sweller, N., & Jones, M. P. (2019). When our hands help us understand: A meta-analysis into the effects of gesture on comprehension. *Psychological Bulletin, 145*, 765–784. https://doi.org/10.1037/bul0000202

Eitel, A., Scheiter, K., & Schüler, A. (2012). The time course of information extraction from instructional diagrams. *Perceptual and Motor Skills, 115*, 677–701. https://doi.org/10.2466/22.23.PMS.115.6.677-701

Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G*power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods, 39*, 175–191. https://doi.org/10.3758/BF03193146

Glaser, M., & Schwan, S. (2015). Explaining pictures: How verbal cues influence processing of pictorial learning material. *Journal of Educational Psychology, 107*, 1006–1018. https://doi.org/10.1037/edu0000044

Gullberg, M., & Holmqvist, K. (2006). What speakers do and what addressees look at. *Pragmatics and Cognition, 14*, 53–82. https://doi.org/10.1075/pc.14.1.05gul

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

Homer, B. D., Plass, J. L., & Blake, L. (2008). The effects of video on cognitive load and social presence in multimedia-learning. *Computers in Human Behavior, 24*, 786–797. https://doi.org/10.1016/j.chb.2007.02.009

Hong, J., Pi, Z., & Yang, J. (2018). Learning declarative and procedural knowledge via video lectures: Cognitive load and learning effectiveness. *Innovations in Education & Teaching International, 55*, 74–81. https://doi.org/10.1080/14703297.2016.1237371

Johnson, M. H., Dziurawiec, S., Ellis, H., & Morton, J. (1991). Newborns' preferential tracking of face-like stimuli and its subsequent decline. *Cognition, 40*, 1–19. https://doi.org/10.1016/0010-0277(91)90045-6

Kalyuga, S., Chandler, P., & Sweller, J. (1999). Managing split-attention and redundancy in multimedia instruction. *Applied Cognitive Psychology, 13*, 351–371. https://doi.org/10.1002/(SICI)1099-0720(199908)13:4<351::AID-ACP589>3.0.CO;2-6

Kizilcec, R. F., Bailenson, J. N., & Gomez, C. J. (2015). The instructor's face in video instruction: Evidence from two large-scale field studies. *Journal of Educational Psychology, 107*, 724–739. https://doi.org/10.1037/edu0000013

Kizilcec, R. F., Papadopoulos, K., & Sritanyaratana, L. (2014). Showing face in video instruction: Effects on information retention, visual attention, and affect. In M. Jones, P. Palanque, A. Schmidt, & T. Grossman (Eds.), *Proceedings of the 32nd annual ACM conference on Human factors in computing systems - CHI '14* (pp. 2095–2102). https://doi.org/10.1145/2556288.2557207. Association for Computing Machinery.

Klepsch, M., Schmitz, F., & Seufert, T. (2017). Development and validation of two instruments measuring intrinsic, extraneous, and germane cognitive load. *Frontiers in Psychology, 8*. https://doi.org/10.3389/fpsyg.2017.01997

Klepsch, M., & Seufert, T. (2021). Making an effort versus experiencing load. *Frontiers in Education, 6*, Article 645284. https://doi.org/10.3389/feduc.2021.645284

Krieglstein, F., Beege, M., Rey, G. D., Sanchez-Stockhammer, C., & Schneider, S. (2023). Development and validation of a theory-based questionnaire to measure different types of cognitive load. *Educational Psychology Review, 35*. https://doi.org/10.1007/s10648-023-09738-0

Lallé, S., Taub, M., Mudrick, N. V., Conati, C., & Azevedo, R. (2017). The impact of student individual differences and visual attention to pedagogical agents during learning with MetaTutor. In E. André, R. Baker, X. Hu, M. M. T. Rodrigo, & B. Du Boulay (Eds.), *Artificial intelligence in education. AIED 2017. Lecture notes in computer science* (Vol. 10331, pp. 149–161). Springer. https://doi.org/10.1007/978-3-319-61425-0_13.

Mayer, R. E. (Ed.). (2014a). *The Cambridge handbook of multimedia learning* (2nd ed.). Cambridge University Press. https://doi.org/10.1017/CBO9781139547369.

Mayer, R. E. (2014b). Principles based on social cues in multimedia learning: Personalization, voice, image and embodiment principles. In R. E. Mayer (Ed.), *The Cambridge handbook of multimedia learning* (2nd ed., pp. 345–368). Cambridge University Press.

Mayer, R. E. (2021). *Multimedia learning* (3rd ed.). Cambridge University Press. https://doi.org/10.1017/9781316941355

Mayer, R. E., & Fiorella, L. (2014). Principles for reducing extraneous processing in multimedia learning: Coherence, signaling, redundancy, spatial contiguity, and temporal contiguity principles. In R. E. Mayer (Ed.), *The Cambridge handbook of multimedia learning* (2nd ed., pp. 279–315). Cambridge University Press.

Meier, J., Jong, B., van Montfort, D. P. de, Verdonschot, A., van Wermeskerken, M., & van Gog, T. (2023). Do social cues in instructional videos affect attention allocation, perceived cognitive load, and learning outcomes under different visual complexity

conditions? *Journal of Computer Assisted Learning, 39*, 1339–1353. https://doi.org/10.1111/jcal.12803
Montoya, A. K., & Hayes, A. F. (2017). Two-condition within-participant statistical mediation analysis: A path-analytic framework. *Psychological Methods, 22*, 6–27. https://doi.org/10.1037/met0000086
Ng, Y. Y., & Przybylek, A. (2021). Instructor presence in video lectures: Preliminary findings from an online experiment. *IEEE Access, 9*, 36485–36499. https://doi.org/10.1109/ACCESS.2021.3058735
Paas, F. G. (1992). Training strategies for attaining transfer of problem-solving skill in statistics: A cognitive-load approach. *Journal of Educational Psychology, 84*, 429–434. https://doi.org/10.1037/0022-0663.84.4.429
Pi, Z., & Hong, J. (2016). Learning process and learning outcomes of video podcasts including the instructor and ppt slides: A Chinese case. *Innovations in Education & Teaching International, 53*, 135–144. https://doi.org/10.1080/14703297.2015.1060133
Pi, Z., Xu, K., Liu, C., & Yang, J. (2020). Instructor presence in video lectures: Eye gaze matters, but not body orientation. *Computers & Education, 144*, Article 103713. https://doi.org/10.1016/j.compedu.2019.103713
Polat, H. (2022). Instructors' presence in instructional videos: A systematic review. *Education and Information Technologies, 28*, 1–33. https://doi.org/10.1007/s10639-022-11532-4
Schramm, H., & Hartmann, T. (2008). The PSI-Process Scales. A new measure to assess the intensity and breadth of parasocial processes. *Communications, 33*, 385–401. https://doi.org/10.1515/COMM.2008.025
Schroeder, N. L., & Cenkci, A. T. (2020). Do measures of cognitive load explain the spatial split-attention principle in multimedia learning environments? A systematic review. *Journal of Educational Psychology, 112*, 254–270. https://doi.org/10.1037/edu0000372
Schüler, A. (2017). Investigating gaze behavior during processing of inconsistent text-picture information: Evidence for text-picture integration. *Learning and Instruction, 49*, 218–231. https://doi.org/10.1016/j.learninstruc.2017.03.001
Schüler, A. (2019). The integration of information in a digital, multi-modal learning environment. *Learning and Instruction, 59*, 76–87. https://doi.org/10.1016/j.learninstruc.2017.12.005
Schüler, A., & Merkt, M. (2021). Investigating text-picture integration in videos with the multimedia contradiction paradigm. *Journal of Computer Assisted Learning, 37*, 718–734. https://doi.org/10.1111/jcal.12518
Sondermann, C., & Merkt, M. (2023a). Like it or learn from it: Effects of talking heads in educational videos. *Computers & Education, 193*, Article 104675. https://doi.org/10.1016/j.compedu.2022.104675
Sondermann, C., & Merkt, M. (2023b). What is the effect of talking heads in educational videos with different types of narrated slides? *Contemporary Educational Psychology, 74*, Article 102207. https://doi.org/10.1016/j.cedpsych.2023.102207.
Stull, A. T., Fiorella, L., & Mayer, R. E. (2021). The case for embodied instruction: The instructor as a source of attentional and social cues in video lectures. *Journal of Educational Psychology, 113*, 1441–1453. https://doi.org/10.1037/edu0000650
Sweller, J. (1999). *Instructional design in technical areas*. ACER Press.
Sweller, J., Ayres, P., & Kalyuga, S. (2011). *Cognitive load theory*. Springer.
Sweller, J., van Merrienboer, J. J., & Paas, F. G. C. (1998). Cognitive architecture and instructional design. *Educational Psychology Review, 10*, 251–296. https://doi.org/10.1023/A:1022193728205
van Gog, T. (2014). The signaling (or cueing) principle in multimedia learning. In R. E. Mayer (Ed.), *The Cambridge handbook of multimedia learning* (2nd ed., pp. 263–278). Cambridge University Press.
van Wermeskerken, M., Ravensbergen, S., & van Gog, T. (2018). Effects of instructor presence in video modeling examples on attention and learning. *Computers in Human Behavior, 89*, 430–438. https://doi.org/10.1016/j.chb.2017.11.038
Wang, J., & Antonenko, P. (2017). Instructor presence in instructional video: Effects on visual attention, recall, and perceived learning. *Computers in Human Behavior, 71*, 79–89. https://doi.org/10.1016/j.chb.2017.01.049
Wang, J., Antonenko, P., & Dawson, K. (2020). Does visual attention to the instructor in online video affect learning and learner perceptions? An eye-tracking analysis. *Computers & Education, 146*, Article 103779. https://doi.org/10.1016/j.compedu.2019.103779
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*, 279–291. https://doi.org/10.1111/mbe.12239
Wang, H., Pi, Z., & Hu, W. (2019). The instructor's gaze guidance in video lectures improves learning. *Journal of Computer Assisted Learning, 35*, 42–50. https://doi.org/10.1111/jcal.12309
Wilson, K. E., Martinez, M., Mills, C., D'Mello, S., Smilek, D., & Risko, E. F. (2018). Instructor presence effect: Liking does not always lead to learning. *Computers & Education, 122*, 205–220. https://doi.org/10.1016/j.compedu.2018.03.011
Zhang, H., Miller, K. F., Sun, X., & Cortina, K. S. (2020). Wandering eyes: Eye movements during mind-wandering in video lectures. *Applied Cognitive Psychology, 34*, 449–464. https://doi.org/10.1002/acp.3632
Zhang, Y., Xu, K., Pi, Z., & Yang, J. (2022). Instructor's position affects learning from video lectures in Chinese context: An eye-tracking study. *Behaviour & Information Technology, 41*, 1988–1997. https://doi.org/10.1080/0144929X.2021.1910731
Zhang, Y., & Yang, J. (2022). Exploring gender differences in the instructor presence effect in video lectures: An eye-tracking study. *Brain Sciences, 12*, Article 946. https://doi.org/10.3390/brainsci12070946
Hooge, I., & Camps, G. (2013). Scan path entropy and arrow plots: Capturing scanning behavior of multiple observers. *Frontiers in Psychology, 4*, Article 996. https://doi.org/10.3389/fpsyg.2013.00996