



A review of eye tracking research on video-based learning

Ruiqi Deng¹ · Yifan Gao²

Received: 20 July 2022 / Accepted: 24 November 2022 / Published online: 7 December 2022

© The Author(s), under exclusive licence to Springer Science+Business Media, LLC, part of Springer Nature 2022

Abstract

Eye tracking technology is increasingly used to understand individuals' non-conscious, moment-to-moment processes during video-based learning. This review evaluated 44 eye tracking studies on video-based learning conducted between 2010 and 2021. Specifically, the review sought to uncover how the utilisation of eye tracking technology has advanced understandings of the mechanisms underlying effective video-based learning and what type of caution should be exercised when interpreting the findings of these studies. Four important findings emerged from the analysis: (1) not all the studies explained the mechanisms underlying effective video-based learning through employing eye tracking technology, and few studies disentangled the complex relationship between eye tracking metrics and cognitive activities these metrics represent; (2) emotional factors potentially serve to explain the processes that facilitate video-based learning, but few studies captured learners' emotional processes or evaluated their affective gains; (3) ecological validity should be improved for eye tracking research on video-based learning through methods such as using eye tracking systems that have high tolerance for head movements, allowing learners to take control of the pacing of the video, and communicating the learning objectives of the video to participants; and (4) boundary conditions, including personal (e.g. age, prior knowledge) and environmental factors (e.g. the topic of videos, type of knowledge), must be considered when interpreting research findings. The findings of this review inspire a number of propositions for designing and interpreting eye tracking research on video-based learning.

Keywords Instructional video · Video-based learning · Eye tracking · Multimedia learning · Review

✉ Yifan Gao
yfgao91@zju.edu.cn

Ruiqi Deng
r.deng@hznu.edu.cn

¹ Jing Hengyi School of Education, Hangzhou Normal University, Hangzhou, China

² College of Civil Engineering and Architecture, Zhejiang University, Hangzhou, China

1 Introduction

Videos are becoming an increasingly important medium of instruction in educational contexts such as MOOCs, flipped classrooms, blended learning, and online training, particularly when teaching and learning are severely interrupted by COVID-19 (Cojean & Jamet, 2022; Deng & Benckendorff, 2021). Research supports that videos offer an advantage to student learning over textbook-style readings when used for pre-class content learning in flipped classrooms at universities (Jensen et al., 2018). In blended learning, university students who watched online instructional videos and participated in classroom teaching achieved better performances than those who only received traditional face-to-face instruction (Expósito et al., 2020). Videos were also found to foster better learning performance than static text and image-based materials in school settings (Chen & Wang, 2011). Researchers have undertaken a series of empirical studies to investigate how learning performance can be strengthened and optimised by designing effective videos (e.g. Arslan-Ari, 2018; Beege et al., 2017; Tarchi et al., 2021). However, these studies have inclined to focus on the impact of design on learning outcomes while *lacking* attention to individuals' learning processes. This in turn results in an insufficient understanding of the mechanisms underlying effective video-based learning (Mayer, 2021).

To explain these underlying mechanisms, research has been conducted on the impact of the design of instructional videos on individuals' learning processes. Self-reporting assessments such as surveys are adopted to make inferences about processes during video-based learning (Boucheix et al., 2018; Fiorella et al., 2017; Hoogerheide et al., 2015; Pan et al., 2020; van Gog et al., 2009). While these measures offer possible explanations for learning activities, they are less successful in documenting temporal fluctuations, non-conscious responses, and latent changes. Consequently, there is a need for measurements that directly capture individuals' moment-to-moment learning processes. The eye tracking method, which has been extensively used by psychologists to investigate information processing during reading (Gordon et al., 2006), has emerged as a promising tool for investigating processes in video-based learning. However, evaluating eye movements is particularly complex for dynamic stimuli such as videos (Madsen et al., 2021). Research has shown that design principles that apply to other formats of learning materials may not work for instructional videos (Pi, Chen, et al., 2020).

Eye tracking data are sometimes used as a proxy for learners' cognitive activities such as paying attention to key elements in a video for further processing and incorporating new information into a coherent cognitive structure (Fiorella et al., 2019). It is argued that 'eye tracking...provides insight into processes underlying learning' (Kok & Jarodzka, 2017, p. 119). In other words, eye tracking technology has the potential to offer effective explanations for how instructional videos facilitate or hamper learning outcomes. Although a growing number of scholars are deploying eye tracking technology in empirical studies (Colliot & Jamet, 2018; van Wermeskerken et al., 2018), no studies have yet addressed

the following research question: How has the utilisation of eye tracking technology and the interpretation of eye tracking metrics advanced our understanding of the mechanisms underlying effective video-based learning? This question is important: if eye tracking research does not provide robust explanations for how instructional videos facilitate or impede learning, then supplemental or alternative measurement strategies must be utilised to scrutinise learning processes. A relevant and equally important research question is: What caution needs to be exercised when interpreting findings of eye tracking research on video-based learning?

This review interrogated the literature to identify key variables (including eye tracking metrics) explored in eye tracking research on video-based learning, and critiqued the evidence for relationships between these variables using Biggs' (1993a) presage-process-product (3P) model as an organisational framework. This work enables researchers and practitioners to better appreciate the pros and cons of utilising eye tracking technology for investigating video-based learning and the intricate relationships among eye tracking metrics, learning processes, and personal and environmental factors. Moreover, the review evaluated where eye tracking research on video-based learning has been concentrated in the global context and where research opportunities for future studies lie.

Video-based learning is defined as a form of learning that enables individuals to acquire knowledge and skills through videos. The term 'video' conveys the same meaning as terms such as 'instructional video', 'educational video', 'video lecture', 'video tutorial', 'video education', and 'video modelling example' used in multimedia learning research. This review defines eye tracking research as evidence-based studies that employ research-grade eye tracking systems and specialised data analysis programmes to examine the positions and movements of an individual's eyes.

2 Method

2.1 Narrative synthesis approach

This review adopted a narrative synthesis approach to summarise, analyse, and explain the data. Narrative synthesis is qualitative and relies primarily on using words and texts to synthesise the findings of multiple studies (Cook et al., 1997). Although narrative synthesis often presents results in a non-numeric manner, the results may take numerical forms. As eye tracking research on video-based learning was heterogeneous for the key variables explored, statistical meta-analysis was not appropriate at this stage. Narrative synthesis produces new insights that reflect the plurality of educational topics (Arai et al., 2007). It was deemed that a narrative synthesis approach would provide a richer and more fine-grained analysis of this topic.

Researchers began narrative synthesis by defining topics and target audiences. The broad topic (i.e. eye tracking research on video-based learning) is of interest to education researchers, academics, teachers, and practitioners such as video producers. This review employed a three-step article research process to identify the most relevant literature through database and manual journal searches. After identifying

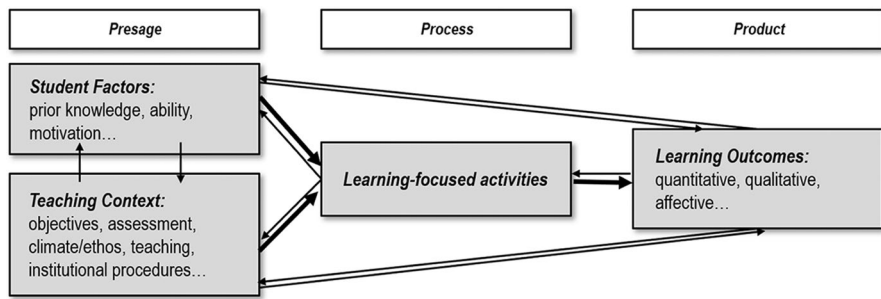


Fig. 1 3P model of teaching and learning. (adapted from Biggs, 1993b)

the relevant literature, the study followed the narrative synthesis practices adopted in internationally recognised academic work (Boyle et al., 2014; Popay et al., 2005). This entailed utilising a conceptual framework, reading and rereading materials, identifying key variables and patterns across studies, and investigating the relationships within and among studies. The present study adopted several tools and techniques to categorise and evaluate articles, including textual descriptions, frequency distributions, tabulations, and groupings. The paper concludes with a discussion of and reflection on these articles.

2.2 Conceptual framework

The 3P model was utilised as an organisational framework to facilitate the evaluation of eye tracking research on video-based learning, to group key variables that are conceptually and empirically similar, and to identify important relationships. Biggs (1993b) elaborates Dunkin and Biddle's (1974) presage-process-product model and conceptualises relationships involving the student, teaching context, student learning processes, and learning outcomes. The model proposes that learning experience comprises three stages: presage, process, and product (Fig. 1). *Presage* factors exist prior to students' engagement in learning. The two types of presage factors were student and teaching context factors. Student factors are relatively stable and learning-related characteristics of the student, such as intellectual abilities, preferred ways of learning, and prior knowledge. Teaching context factors are contextual and teaching-related factors, such as course structure, assessment methods, and institutional climate. These two sets of presage factors interact at the *process* level and produce a particular teaching and learning mix that determines learning-focused activities (Biggs, 1993b). The interaction of these factors determines the *product* of learning, which can be described and evaluated quantitatively, qualitatively, and affectively. The heavy arrows in Fig. 1 indicate the main directional flow, and the flow between key factors is regarded as bidirectional (Biggs, 1993a).

The 3P model was adopted for this review because it heuristically outlines important learning-related variables and their relationships from presage to process to product and provides scope for adaption to the context of video-based learning. Biggs (2003) indicated that the 3P model could be adapted to different learning

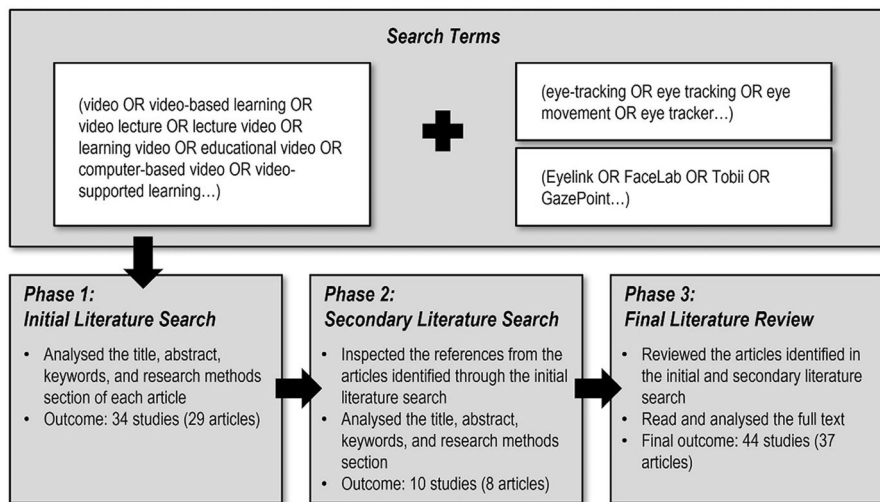


Fig. 2 Article search strategy

environments, and any identifiable factors that affect learning can be accommodated in this model. In this review, student factors describe student characteristics that may affect video-based learning, such as social demographics, prior knowledge of video content, and prior interest in video content. The teaching context factors describe the characteristics of videos, such as video topic and length. Student and teaching context factors interact and codetermine learning-related activities at the process level, such as attention and engagement. The interaction among these factors determines the outcomes of video-based learning, such as academic performance and satisfaction.

2.3 Article search strategy

The researchers used two sets of search terms in six electronic academic databases: ScienceDirect, Springer, Taylor and Francis, Web of Science, Wiley, and Google Scholar. The first set narrowed the scope to studies that focused on video-based learning, and the second set narrowed the scope to studies that adopted eye tracking technology. For the first set, the researchers used search terms such as video lecture, instructional video, and educational video. For the second set, the researchers adopted both general terms (e.g. eye tracking, eye tracking, eye movement) and specific brands representing the application of eye tracking technology (e.g. Eye-link, FaceLab, Tobii). The search terms were identified through exploratory search results; consultation with experts in the field of educational technology; and a review of titles, abstracts, and keywords of the identified articles. Examples of search terms are shown in Fig. 2. The operator ‘OR’ was used to separate the search terms within each set, and the operator ‘AND’ was used to combine the search terms in the two sets.

The article search process followed procedures implemented in previous review studies (Deng et al., 2019, 2022) and contained three steps. In the first phase, the researchers examined the titles, abstracts, keywords, and research methods to determine whether an article should be included. Articles meeting the predetermined selection criteria were identified (Table 1). Articles which did not employ eye tracking technology to investigate video-based learning were not eligible for review (e.g. Fiorella et al., 2017). Nevertheless, these articles were used as background literature to help interpret the findings. Research-grade eye tracking systems often guarantee a sampling rate higher than 50 Hz and support specialised data analysis programmes. Articles which adopted off-the-shelf webcams to record learners' eye movements and employed humans to assess and code gaze from video recordings were also excluded due to low sampling rates (e.g. Phillips et al., 2016). In the second phase, references from articles identified through the initial literature search were inspected to identify potentially relevant papers that had been overlooked during the database search. Duplicated articles and those not meeting the selection criteria were eliminated from the pool. In the third phase, the researchers read and evaluated the full texts of the articles identified in the first and second phases to ensure that they addressed the research question and met the selection criteria. At the end of the selection process, the combined results from the database and manual journal search yielded 37 articles. As some articles contained more than one study, 44 studies were included in the analysis.

2.4 Analysis

Each article was treated as a basic unit of analysis. An overall synthesis table was produced after reading all the articles. The table contains descriptive information such as authors, year of publication, journal name, title, location of the study, research objectives, and major findings. Drawing on the 3P model, the key variables investigated and reported were extracted from all the articles and recorded in the table. These variables were classified into one or more of four categories: student factors, teaching context factors, learning-focused activities, and learning outcomes. Each category comprised multiple subcategories. For example, the learning outcomes category contained subcategories such as knowledge retention, knowledge comprehension, knowledge transfer, satisfaction, and perceived learning. Wang et al. (2020d) examined the impact of cues on learning processes and outcomes, and the outcomes were measured using retention and transfer tests. Key information about retention and transfer tests (e.g. question types and materials) were documented in the learning outcomes category and the knowledge retention and transfer subcategories of the table. This categorisation facilitated the identification of similarities and differences between articles. The labels of the subcategories were constantly refined and merged to better capture similarities and differences.

Patterns and themes were identified across the selected articles using an iterative rereading process. This process ensured that all important information was identified and analysed. The following section summarises and compares evidence related to the four categories in the adapted 3P model. To ensure an informative and concise

Table 1 Selection criteria for articles

Selection criteria		Rationale
Topic of interest	Video-based learning	The paper is set to review eye tracking research on video-based learning.
Research method	Eye tracking technology	
Type of scientific research	Empirical studies	This review is based on knowledge derived from experiences and observations, rather than personal opinions or assumptions. Conceptual work, editorials, and critiques of the literature are not included.
Type of publications	Peer-reviewed academic journal articles	Non-academic sources (e.g. news, webpages) may not be scientific. Non-peer-reviewed academic sources (e.g. technical reports) may not be rigorous. Conference papers sometimes use a very small sample size (e.g. Altan & Cagiltay 2015) and lack adequate details required to complete this review.
Quality of publications	Transparent data collection procedures and analysis methods	Sources that are not methodologically sound yield untenable conclusions and would affect the quality of this review. It was expected that procedural details such as participant recruitment, materials, apparatus, and measurements are clearly explained in the sources. Restrictions were not placed on the participants' educational stages.
Date of publications	Between 2010 and 2021	Both older and newer eye tracking technology provides insight into video-based learning. As few academic sources employed eye tracking technology to investigate video-based learning before 2010, this review set the date range for publications to 2010–2021.
Language	English	It can be difficult to access and evaluate the quality of sources published in languages other than English.

Table 2 Journal information about the reviewed studies

Journal	Frequency	Percentage
Computers in Human Behaviour	7	15.91
Learning and Instruction	6	13.64
Computers & Education	5	11.36
Journal of Educational Psychology	3	6.82
The Proceedings of the National Academy of Sciences	3	6.82
Others	20	45.45

review, examples of reviewed articles were selected and displayed in tables to complement the textual descriptions. Journal information regarding the reviewed studies is presented in Table 2.

3 Results

3.1 Student factors

All the reviewed studies reported the number, sex, age, and occupation of the participants. The number of participants varied from 21 (Jarodzka et al., 2010) to 174 (Pi et al., 2020b). The mean number of participants in the study was 68. The highest male-to-female ratio was 3:1 (Pi et al., 2017), and the highest female-to-male ratio was 8:1 (Pi et al., 2020a). The average age of the participants ranged from 12.1 (van Marlen et al., 2018) to 30.7 (Gegenfurtner et al., 2017). Age or occupation was not controlled for in eye tracking research on video-based learning because most of the studies were conducted with university students ($n=42$). Only two studies recruited secondary school students (van Marlen et al., 2018) and high school students (Wang et al., 2016) as participants. No reviewed studies recruited participants from primary schools.

Thirty-two studies considered prior knowledge to be an important student factor. A variety of assessment methods, such as true or false (Pi et al., 2020a), multiple choice (Montero Perez et al., 2015), fill-in-the-blank (Pi et al., 2022), open questions (van Marlen et al., 2018), and Likert scales (Wang et al., 2020b), were used to measure participants' prior knowledge of the video content. All 32 studies employed objective tests to assess the participants' prior knowledge, with two exceptions. Gegenfurtner et al. (2017) accepted participants' own account of prior knowledge levels, and Wang et al. (2020b) used a subjective test by asking participants to self-report their prior knowledge of video content. Such efforts served to increase the likelihood that different groups had a similar level of prior knowledge but could not guarantee equality. To ensure that prior knowledge did not exert undue influence on learning outcomes, the reviewed studies either operationalised prior knowledge as covariates (Wang et al., 2020a) or established that different groups did not differ in prior knowledge scores (Wang et al., 2020d). Video design interventions that

work for people with higher domain knowledge may be ineffective for individuals who have lower levels of prior knowledge (Gegenfurtner et al., 2017), and vice versa (Krebs et al., 2019; van Marlen et al., 2018). Experimental research that does not assess or control for prior knowledge (e.g. Ouwehand et al., 2015) may potentially bias or attenuate the effect of the design intervention on learning outcomes.

3.2 Teaching context

The topic and length of videos have been descriptively reported in all the reviewed studies. The topic of videos differs widely among the reviewed studies and demonstrate no distinct patterns, ranging from education (Pi et al., 2019), psychology (Kruger & Steyn, 2014), and management (Wang et al., 2016), to programming (Kokoç et al., 2020), statistics (Wang et al., 2020), and mathematics (van Marlen et al., 2018). Of the 44 studies, 35 retained the length of the video for < 10 min. The length of the videos was up to 44 min (Kruger & Steyn, 2014). One study did not disclose the length of the videos used in their study (Wang et al., 2019b). To ensure the accuracy of the eye tracking data, eight studies clearly indicated that a chinrest was used to minimise head movement, and twelve studies acknowledged that participants were not allowed to pause or rewind while watching an instructional video.

The teaching context can also be interpreted from the perspective of being manipulated and operationalised as independent variables in a controlled experiment (Table 3). This review classified these studies into three categories. The first research category explores the effects of an instructor's presence in videos and accompanying social cues on student learning. Research supports that videos featuring both the instructor and content enhance learning performance (Colliot & Jamet, 2018; Pi & Hong, 2016; van Gog et al., 2014) for both easy (Wang & Antonenko, 2017) and difficult topics (Wang et al., 2020b). This design principle was tested with videos on the topics such as attachment, Ebola, sleep, and mathematics, with a length of between 3 and 10 min. Eye tracking data indicate that the instructor-present videos resulted in more fixation counts, longer dwell time (Pi & Hong, 2016) and a higher percentage of fixation on the instructor (Wang & Antonenko, 2017), which suggests that the processing of the instructor's image may have provided social cues to facilitate the processing of cognitively relevant information and elicit beneficial social-emotional responses from learners (Wang et al., 2020b). That is, eye tracking metrics may have served as process factors, providing a way to explain the mechanism by which instructor presence affects learning performance.

However, the effects of instructor presence on learning remain inconclusive. van Wermeskerken and van Gog (2017) and van Wermeskerken et al. (2018), for instance, found that the presence of an instructor had neither beneficial nor detrimental effects on learning performance. Even though instructor presence did not affect learning performance, it still increased the percentage of dwell time on the instructor (van Wermeskerken & van Gog, 2017). This observation suggests that while eye tracking metrics can reveal learners' allocation of visual attention, these metrics cannot always explain why learning is facilitated or not. This is likely because learners' allocation of visual attention conveys multiple cognitive meanings

Table 3 Factors tested in controlled experiments

Author	Tested factor	Experimental conditions	Key finding
de Koning et al. (2010)	Non-social cue	Single cue, multiple cues, or no cue	Watching a video with a single spotlight cue or with multiple spotlight cues did not lead to better learning performance than viewing the video without cues
Jarodzka et al. (2012)	Non-social cue	Circle, spotlight, or no cue	Spotlight cues enhanced interpretation performance
Jarodzka et al. (2013)	Non-social cue	Dot, spotlight, or no cue	Cues enhanced the perceptual task performance
Jamet (2014)	Non-social cue	Colour change, no colour change	Colour change enhanced the retention of signalled information
van Gog et al. (2014)	Instructor's presence	Instructor-present or instructor-absent	Instructor presence enhanced learners' performance
Montero Perez et al. (2015)	Non-social cue	Full or keyword captioning	Keyword captioning promoted learners' performance in form recognition tests
	Motivational strategy	Informed or not informed of a pending test	Informing learners about an upcoming vocabulary test promoted learners' performance in meaning recall tests
Ouwehand et al. (2015)	Social cue	Gaze cue, gesture and gaze cue, or no cue	The instructors' gaze or gesture had no effects on transfer performance
van Marlen et al. (2016)	Non-social cue	Meaningful cues, meaningless cues, or no cue	Visual cues had no effects on learning performance
Pi and Hong (2016)	Non-social cue	Cues or no cue	
	Instructor's presence	Slides, instructor, or slides and instructor	Videos showing both slides and the instructor enhanced learners' performance
van Wermeskerken and van Gog (2017)	Instructor's presence, social cue	No face visible, face visible with gaze guidance, or face visible without gaze guidance	Instructor presence neither facilitated nor hampered retention and transfer performance

Table 3 (continued)

Author	Tested factor	Experimental conditions	Key finding
Wang and Antonenko (2017)	Instructor's presence	Instructor-present or instructor-absent	Instructor presence improved retention performance for the easy topic, reduced cognitive load for the difficult topic, and enhanced perceived learning and satisfaction for both easy and difficult topics
Colliot and Jamet (2018)	Instructor's presence	Instructor-present or instructor-absent	Instructor presence enhanced retention performance
Stull et al. (2018)	Social cue	Conventional or transparent whiteboard	There was no difference in learning performance between individuals who watched a transparent whiteboard video and those who viewed a conventional whiteboard video
Wang et al. (2018)	Instructor's presence Social cue	Embodied pedagogical agent (PA) or no PA High or low embodiment	A highly embodied PA enhanced learners' performance.
Krebs et al. (2019)	Motivational strategy	No guidance, guidance of a successful learner, or guidance of a peer	The framing of the instructor as a peer learner enhanced the comprehension of learners with lower domain knowledge.
Pi et al. (2019)	Social cue	Pointing, depictive, or no gesture	The instructor's gestures had a more positive effect for learners with low prior knowledge
Wang et al. (2019a)	Social cue	With or without gaze guidance	The instructor's gaze guidance improved learners' performance
Chisari et al. (2020)	Non-social cue	Translucent blue dots or no cue	Cues did not lead to higher learning outcomes
Pi et al. (2022)	Motivational strategy	Happy or neutral face	The instructor's happy facial expression enhanced learners' performance
	Motivational strategy	Direct or averted gaze	There was no significant effect of gaze on learners' performance

Table 3 (continued)

Author	Tested factor	Experimental conditions	Key finding
Wang et al. (2020b)	Social cue	Instructor-present or instructor-absent	Instructor presence improved transfer performance and reduced cognitive load for a difficult topic, and enhanced satisfaction for both easy and difficult topics
Pi et al. (2020b)	Social cue	Direct, guided, or averted	The instructor's guided gaze improved retention and transfer performance
	Motivational strategy	Frontal or lateral	There was no significant effect of body orientation on learners' performance
Wang et al. (2020d)	Non-social cue	Textual, visual, textual and visual, or no cue	The visual cues and combined textual and visual cues improved both retention and transfer performance
Pi et al. (2021)	Social cue	Direct or guided gaze	Videos showing an instructor's guided gaze and surprised face resulted in lower learning performance
	Motivational strategy	Surprised or neutral face	
Zhang et al. (2022)	Instructor's presence, social cue	Instructor on the left, instructor in the middle, instructor on the right, or no instructor	Videos presenting the instructor on the right side of the screen improved learners' performance and satisfaction

that can be difficult to interpret solely relying on the analysis of global eye movement measures such as dwell time.

When the instructor's image is presented in videos, the influence of accompanying social cues on student learning, such as the instructor's eye gaze and gestures, was further investigated. It is contended that the activation of social schemata triggered by social cues leads to (para-)social processes influencing all cognitive processes in multimedia learning environments (Schneider et al., 2022). Research indicates that instructors' guided gaze promotes learners' performance (Pi et al., 2020b). Similarly, a pedagogical agent using a handheld pointer that signals where to look on the screen also enhances learning performance (Wang et al., 2018). Eye tracking data show that guided gaze (Pi et al., 2020b) and using a pointer (Wang et al., 2018) contribute to a longer fixation time on the learning content. A plausible explanation is that social cues effectively draw learners' attention to important, relevant materials for cognitive processing, thereby improving students' learning performance. However, the effect of social cues on student learning is not always positive. Although the main effect of guided gaze on learning was reported to be significant (Pi et al., 2020b), a combination of guided gaze and a surprised face was found to decrease learners' dwell time on the learning content and subsequently their performance (Pi et al., 2021). These findings highlight the importance of considering the possible interaction effects between embedding social cues and applying motivational strategies in instructional videos, which instead of benefitting, may hamper learning processes and outcomes.

Eye tracking metrics shed light on visual attention distribution that explains why learning performance was facilitated or hindered. However, eye tracking metrics alone may not show whether individuals who spent more time looking at task-relevant areas successfully processed this information and that this would facilitate learning. The combination of two or more social cues, such as the instructor's guided gaze and pointing gestures, was found to effectively direct learners' visual attention to task-relevant areas but had no effect on learning performance (Ouweland et al., 2015), highlighting the necessity to perform a more nuanced analysis of eye movement indicators and/or using additional measures to comprehend the meaning behind visual attention distribution.

The second category of research explores the effects of non-social cues, such as video captions, textual cues, and visual cues, on learning performance. For instance, Montero Perez et al. (2015) found that learners who watched videos showing keyword captions outperformed their counterparts who watched videos with full captions in a retention test. Eye tracking data demonstrated that the former group had a longer total fixation duration for target words. These findings signal that the relationship between captioning and retention performance is mediated by the allocation of visual attention. However, it is worth noting that Montero Perez et al.'s (2015) study was undertaken in the context of second language acquisition, where learning novel words was given a high priority, and when learners were informed of an upcoming test. The design principle may not be generally applicable. A longer total fixation duration contains multiple layers of cognitive significance, such as increased intention to acquire knowledge and processing problems. To accurately interpret allocation of visual attention, additional eye movement indicators (e.g. second pass

time) and complementary measures (e.g. think-aloud protocols) (Gegenfurtner et al., 2017) are worth consideration.

Eye movement indicators revealed visual attention processes that helped to explain why learning performance was impacted. For example, Wang et al. (2020d) found that visual cues were more effective in prompting learners' performance than no cues, whereas de Koning et al. (2010) did not observe such a difference. A possible explanation is that Wang et al. (2020d) designed dynamic lines, highlighted colouring, and moving dots to serve as visual cues, which helped learners pinpoint the exact location of key information. In comparison, de Koning et al. (2010) used spotlight cues, which may have limited the size of the content area learners focused on, but failed to guide processing cognitively relevant information within this area. This conjecture is substantiated by eye tracking data: visual cues comprising lines, colouring, and dots reduced the time of the first fixation to the cued areas (Wang et al., 2020d), whereas in de Koning et al.'s (2010) study spotlight cues increased attention but not necessarily to cued areas. Jarodzka et al. (2012), by contrast, found that spotlight cues enhanced learning performance. The varying effectiveness of spotlight cues can be explained by eye tracking metrics showing that spotlight cues in Jarodzka et al.'s (2012) research prompted students to look significantly earlier, not longer, at all relevant cued parts. This result is in line with Chisari et al.'s (2020) finding that translucent blue dots helped learners to look faster at referenced information, thereby improving learning outcomes. Despite the inconsistency in study results and the type of visual cues employed, eye tracking metrics provided a way to explain how timely selection of the right information mediates the effect of non-social cues on learning performance.

The last research category explores the effects of motivational strategies on student learning. Montero Perez et al. (2015), for example, revealed a significant positive effect on retention performance by announcing an upcoming test to students. Specifically, eye tracking data showed that individuals who received a test announcement showed longer second pass reading times, indicative of the reanalysis of target words, outperformed those who were not informed. This is aligned with Madsen et al.'s (2021) finding that individuals in the incidental learning condition had lower attentional levels than those in the intentional learning condition, resulting in less correlated eye movements across learners. These results indicate that motivational factors may play a prominent role in video-based learning.

However, eye tracking metrics alone may not always explain how motivational strategies mediate learning. For instance, research shows that instructors with happy faces¹ improved learning performance (Pi et al., 2022). Despite the improvement, eye tracking data revealed that a happy face did not lead to a longer dwell time on the content or the instructor area (Pi et al., 2022). This observation suggests that factors beyond the allocation of visual attention may have mediated the learning process. In other words, researchers may need to consider factors beyond the allocation

¹ The instructor having a happy face is considered a motivational strategy instead of a social cue. This is because happy faces do not have a directive function. That is, they do not direct learners' visual attention to task-relevant areas at the right time.

of visual attention when identifying the principles of effective video design and use alternative measures to capture individuals' emotional and motivational processes during video-based learning.

3.3 Learning-focused activities

Eye tracking data were used in all the reviewed studies to investigate the learning processes. Prior to analysing the eye tracking data, one or more areas of interest (AOIs) were predetermined by researchers to select specific regions of the video material and extract metrics specifically for those regions. The selected AOIs differed among the reviewed studies, but common AOIs were the instructor area, social cues, content area, non-social cues, progress bar, caption area, blank area of the screen, and entire screen.

Table 4 illustrates eight eye tracking metrics used more than once in the reviewed studies to capture the learning processes and the definition of each metric. The most frequently used metric is *dwelt time*, which depicts the total amount of time spent looking at an AOI. This metric is strongly correlated with *the fixation count* (Tullis & Albert, 2013), which is measured by counting the number of fixations on an AOI. This correlation explains why many scholars report either dwell time or fixation count, but not both (e.g. Stull et al., 2018). The reviewed studies used dwell time more frequently than the fixation count. This is likely because instructional videos contain dynamic content, and eye movements such as 'smooth pursuit' cannot be appropriately captured by using the number of fixations. An increased dwell time and fixation count can imply complexity, engagement, or interest (Geisen & Romano Bergstrom, 2017). *Percentage of dwell time* and *percentage of fixations* serve similar purposes, except that they capture relative rather than absolute attention allocation.

Fixation transitions, which describe the frequency of transitions between AOIs, were also repeatedly used in the reviewed studies. Fixation transitions can be used to infer learners' attempts to establish connections between pieces of information or challenges encountered by learners when coordinating multimedia elements (Alemdag & Cagiltay, 2018). Additionally, *average fixation duration* indicates how long the average fixation lasted, and a longer duration often indicates that individuals spend more time analysing the content or expend more effort to solve the task (Sharafi et al., 2015). *Time to first fixation* represents the amount of time taken to first pay attention to specific AOIs in a video scene. It can provide information about learners' visual search speed or how certain aspects of the scene are prioritised (Neta et al., 2017). *Fixation dispersion* represents how fixations are spread across a scene, and can be used to signify an internal deviation in the content of thoughts from ongoing tasks (Faber et al., 2020).

Most reviewed studies used eye tracking metrics as a proxy for the allocation of visual attention but under various terms, including but not limited to: 'attentional resources' (Montero Perez et al., 2015, p. 323) 'attentional bias' (Ouwehand et al., 2015, p. 85), 'amount of attention' (van Wermeskerken & van Gog, 2017, p. 100), 'attentional dynamics' (Wang & Antonenko, 2017, p. 87), 'attention allocation' (Pi, Xu, et al., 2020, p. 5), 'visual attention distribution' (Wang et al., 2020b, p. 2), and

Table 4 Repeatedly used eye tracking metrics and their definitions

Metric name	Definition	Number of the reviewed studies using this metric	Example study
Dwell time	The total amount of time that an individual spent looking within an AOI.	27	Kokoç et al. (2020)
Fixation count	The frequency of fixations on an AOI in a period of time.	11	Wang and Anttonen (2017)
Percentage of fixations	The number of fixations on a given AOI divided by the number of fixations on all AOIs, or the number of total fixations on the video.	8	van Gog et al. (2014)
Percentage of dwell time	The total dwell time on a given AOI divided by the fixation duration on all AOIs, or the total fixation duration on the video.	7	Zhang et al. (2022)
Average fixation duration	The average time for fixations on an AOI.	7	Colliot and Jamet (2018)
Time to first fixation	The amount of time it takes for an individual to look at an AOI for the first time from stimulus onset.	7	Wang et al. (2020d)
Fixation transitions	The number of fixation transitions an individual made between two or more AOIs.	7	Wang et al. (2020b)
Fixation dispersion	Dispersion on the screen divided by the maximum dispersion.	3	Jang et al. (2020)

‘attentional state’ (Madsen et al., 2021, p. 1). The reviewed studies provide evidence that eye tracking technology can identify what information is visually attended to by the learner and for how long, thereby detecting attention during video-based learning. However, many scholars have not only considered eye tracking metrics as an indication of the allocation of visual attention, but also learners’ cognitive processes (e.g. Pi & Hong, 2016; Wang et al., 2019b). This is based on the eye-mind assumption, which posits that there is no appreciable delay between what is being fixated on and what is being processed by the learner (Just & Carpenter, 1980).

Researchers commonly interpret the cognitive meaning of the same eye tracking metric differently. For instance, Krebs et al. (2019) and Wang et al. (2020) interpreted the number of fixation transitions between two AOIs as the cognitive process of organising and integrating information that would benefit learning, whereas Wang et al. (2020b) considered the same measure as the amount of split attention that could harm learning. While an increase in fixation count may imply that an AOI is prominent (Kokoç et al., 2020), the same metric can also represent a higher level of difficulty when processing an AOI (Wang et al., 2019b). A longer fixation duration on an AOI can evince the investment of more cognitive resources (Pi et al., 2020a), but it can also be used to approximate mind-wandering (Jang et al., 2020).

These discrepancies are not surprising, given that interpretation of eye tracking metrics often depends on the context of each study. However, researchers have sometimes failed to provide strong justifications as to why the eye tracking metrics they used can represent cognitive processes they assumed to reflect. For instance, Wang et al. (2020b, p. 6) utilised ‘the number of transitions between the two [AOIs] to understand the amount of split attention’ without explaining why this metric represented split attention, not information integration. Wang et al. (2020d, p. 6) considered ‘transitions between...AOIs as an indication that represented attempts at organising and integrating information’ because ‘in previous studies, transition measures were used to represent attempts at organising and integrating information’. Similarly, Krebs et al. (2019, p. 132) interpreted the same metric as an indicator of the learners’ attempts to integrate information, which was ‘based on previous research’. Deducing cognitive processes directly from eye movement indicators without elaborating on why the used indicators can represent the cognitive processes runs the risk of confabulating the cognitive meaning of eye tracking metrics and deriving erroneous working mechanisms. The actual cognitive meaning behind these eye tracking metrics awaits further clarification, empirical validation, and methodological triangulation.

This review shows that eye tracking metrics provide a way to explain the mechanisms underlying effective video-based learning that otherwise would be difficult to discover through traditional measurement approaches. For instance, Pi et al. (2020b) revealed that learners watching instructional videos with a guided gaze showed an improved learning performance against individuals viewing the video with a direct or averted gaze; here, the guided gaze encouraged learners to pay more attention to important, task-relevant areas. Despite its potential usefulness, not all the reviewed studies successfully explained the mechanisms underlying effective video-based learning through utilising eye tracking technology. Zhang et al. (2022), for example, found that learners achieved better performance when the instructor appeared

on the right side of the instructional video compared to those without an instructor on screen; however, eye tracking metrics provided no evidence that having the instructor on the right side prompted learners to pay more attention to the learning content or make more meaningful transitions between the learning content and the instructor. Wang et al. (2020b) found that instructor presence positively influenced learning performance for the difficult topic video yet did not affect learning performance for the easy topic video; however, eye tracking metrics indicated that the instructor attracted more fixations and a longer dwell time in both difficult and easy topic videos. In other words, analysing eye tracking metrics alone cannot explicate the mechanism by which instructor presence enhances learning performance (or lack thereof) in videos with varying levels of difficulty.

In addition to the eye tracking approach, self-report measures were adopted to explore the learners' experiences and complement eye movement measures. Table 5 displays the four psychological constructs used more than once in the reviewed studies and their definitions. The most frequently measured construct is *cognitive load*, which in this context evaluates the mental resources learners perceive to have used in working memory to comprehend the content of instructional videos. Wang and Antonenko (2017) and Wang et al. (2020b), for example, revealed that instructor presence reduced cognitive load when students were learning from instructional videos. That is, self-report measures explained how the effect of instructor presence in videos on learning performance was mediated by the reduction in cognitive load. Furthermore, the positive effect of the instructor appearing on the right side of the screen on learning performance was explained by students' self-report motivation, not eye tracking metrics (Zhang et al., 2022). Although the discrete nature of self-report data implies that the temporality of learners' non-conscious processes is ignored, they serve to capture individuals' perceptions of the learning process and exist as an alternative avenue to explain why a video design intervention improved the learning outcome. Self-report measures also help rule out mechanisms that do not apply. For example, the effects of the instructor's face (Colliot & Jamet, 2018) and facial expression (Pi et al., 2021) on learning performance cannot be explained by the medium of learners' subjective ratings of social presence. As such, self-report measures should not be dismissed when investigating learning-focused activities during video-based learning.

The reviewed studies also assessed learners' social presence, engagement, and situational interest using a self-reporting approach. Social presence, engagement, and situational interest are distinct sets of psychological constructs; however, they all touch the affective or emotional aspects of learning to varying degrees. For instance, Colliot and Jamet (2018, p. 1423) measured social presence based on a semantic differential scale 'cold-warm'; Zhang et al. (2020, p. 452) assessed engagement on a Likert scale by asking learners to indicate the degree to which 'the material covered...was interesting'; and Wang et al. (2020b, p. 146) evaluated situational interest on a Likert scale and asked respondents to rate the level of agreement with the statement 'I am willing to watch more videos like this because it is exciting...'. This observation highlights that emotion could be a salient factor that mediates learning processes in video-based learning. Despite its potential usefulness, only seven studies attempted to capture learners' self-reported emotional state. Measurement

Table 5 Constructs and their definitions

Construct	Definition	Number of the reviewed studies measuring this construct	Example study
Cognitive load	'the load that performing a particular task imposes on the cognitive system' (Sweller et al., 1998, p. 266).	6	Wang and Antonenko (2017)
Social presence	'the degree of salience of the other person in the interaction and the consequent salience of interpersonal relationships' (Short et al., 1976, p. 65).	3	Colliot and Jamet (2018)
Engagement	'a meta-construct that includes behavioural, emotional, and cognitive engagement' (Fredricks & McColskey, 2012, p. 764).	2	Zhang et al. (2020)
Situational interest	'temporary interest that arises spontaneously due to environmental factors such as task instructions or an engaging text' (Schraw et al., 2001, p. 211).	2	Wang et al. (2020b)

of other affective outcomes, such as change in affect (Wong & Adesope, 2021), was not observed in the reviewed studies. None applied physiopsychological measures to track changes in learners' continuous emotion during video-based learning.

3.4 Learning outcomes

This review categorises learning outcomes based on knowledge tests and self-report measures. Knowledge tests are considered more objective than self-report measures and are used to determine whether participants have shown discipline-specific cognitive learning gains after viewing an instructional video. This review further divided knowledge tests into retention, comprehension, and transfer tests. Retention tests are a form of testing that estimates learners' memory of the material and normally involves the task practiced during an acquisitional phase (Seel, 2012); comprehension tests assess learners' ability to read and mentally grasp the meaning of information (Conradie & Frith, 2000); and transfer tests measure the transferability of what was learned in practice conditions to a novel situation (Seel, 2012). It is generally accepted that knowledge retention is less cognitively complex than knowledge comprehension, which in turn is less intricate than knowledge transfer (Krathwohl, 2002). This review showed that knowledge acquisition was assessed at all three levels of retention ($n=27$), comprehension ($n=15$), and transfer ($n=26$).

Of the 44 studies, 41 used knowledge test scores as learning outcome indicators. Twelve studies adopted a single test type to assess knowledge retention, comprehension, or transfer. Twenty-nine studies combined two test types to measure knowledge retention and transfer (e.g. Wang et al., 2020b), knowledge retention and comprehension (e.g. Stull et al., 2018), and knowledge comprehension and transfer (e.g. Pi & Hong, 2016). However, these studies tended not to disclose the cognitive objective that the participants were expected to achieve after watching a video. Instructional videos are produced to help learners perform tasks at various levels of cognitive complexity, ranging from less cognitively complex tasks, such as recalling previously learned information to drawing out factual answers (Xiang & Miller, 2020), to more cognitively complex tasks, such as applying previously learned knowledge on new scenarios (Garrett, 2021). Transparent reporting of the educational objectives of a video and the rationale for adopting a certain type of knowledge test can contribute to clarity and validity and assist the reader in understanding the cognitive level at which video design principles may apply.

Compared to objective testing, fewer studies have adopted self-report measures for evaluating learning outcomes. Self-report measures were either used alone (J. Wang et al., 2019) or in conjunction with knowledge tests (Stull et al., 2018). This review divided self-report measures into learner satisfaction ($n=5$), perceptions of the instructor ($n=2$), and perceived learning ($n=2$). Learner satisfaction is derived by asking individuals to rate the overall quality of their educational experience (Benton & Cashin, 2014); perceptions of the instructor provide information regarding learners' subjective feelings towards the instructor (Harnish & Bridges, 2011); and perceived learning represents changes in people's perceptions of knowledge and skills before and after the learning experience (Calvo-Ferrer, 2017).

Perceived learning was operationalised as an outcome indicator to assess cognitive gain. Learner satisfaction and perceptions of the instructor, on the other hand, were operationalised as learning outcome indicators to assess participants' affective gains, which are viewed by some educators as an equally important educational outcome (Rogaten et al., 2019). Evaluating affective outcomes could be particularly important in research exploring the effectiveness of design interventions orchestrated to arouse learners' social-emotional responses (e.g. Wang et al., 2018), which are also likely to affect cognitive processes during video-based learning. This effort may help identify factors that facilitate or impede video-based learning from emotional and motivational perspectives.

4 Discussion

This review adopted a narrative synthesis approach to analyse and critique 44 eye tracking studies on video-based learning. The review moves the field forward by making four important contributions.

First, it shows that not all the studies managed to explain the mechanisms underlying effective video-based learning through employing eye tracking technology. The instructor's image, for instance, attracted a substantial amount of visual attention from learners irrespective of whether their learning performance was improved (e.g. Colliot & Jamet, 2018) or not (e.g. van Wermeskerken & van Gog, 2017). Such ambiguity highlights the necessity of introducing additional measures to explain the underlying mechanisms. The review identified that self-report methods exist as an alternative avenue to explain why a video design intervention could enhance the learning outcome when eye tracking technology fails to fulfil the task (e.g. Zhang et al., 2022). Variables such as cognitive load (e.g. Wang & Antonenko, 2017) have been measured through self-report methods and used in association with eye tracking data to probe the learning process and explain the working mechanisms of instructional videos. However, adopting self-report measures to capture the learning process is subject to measurement problems such as social desirability bias (Pi et al., 2017) and floor effect (Wang et al., 2020b), which may attenuate the effect of interventions on learning outcomes. These potential weaknesses call for non-invasive physiopsychological measures to complement eye tracking and self-report data when investigating the mechanisms underlying effective video-based learning, such as utilising electroencephalography to continuously assess learners' cognitive load (Wang et al., 2020c).

In addition, this review shows that few studies disentangled the complex relationship between eye tracking metrics and the cognitive activities these metrics represent. Researchers have challenged the eye-mind assumption through empirical investigations, maintaining that the interpretation of eye tracking parameters should depend on the context in which they are applied, and under certain circumstances, these parameters do not align with learners' cognitive processes (Faber et al., 2020; Schindler & Lilienthal, 2019; Wu & Liu, 2022). Although several reviewed studies asserted that they were firmly based on the eye-mind assumption, few have disentangled the complex relationship between eye tracking metrics and cognitive activities

these metrics truly represent; rather, they come up with ad hoc explanations for the observed set of eye movement indicators. Eye movement indicators reflect ‘ongoing processes to the extent that the processes depend on the encoding of information’ (Anderson et al., 2004, p. 230). Because these indicators can mirror the combined effects of several ongoing cognitive processes (Anmarkrud et al., 2019; Kok & Jarodzka, 2017), deciphering the cognitive meaning of learners’ visual attention distribution based solely on eye tracking technology can be difficult. Future research could triangulate eye tracking data with additional measurement approaches, such as cued retrospective reporting (Bender et al., 2021), to minimise ambiguity and avoid misinterpretation of eye movement indicators.

Second, this review found that emotional factors can potentially explain the processes that facilitate video-based learning, yet few studies captured learners’ emotional processes and evaluated their affective gains. Motivational strategies, such as instructors showing a happy face in videos (Pi, Chen, et al., 2020), did not influence learners’ allocation of visual attention but still improved learning performance. Informing students about upcoming tests (Montero Perez et al., 2015), for example, motivated learners to invest additional cognitive resources when watching instructional videos. This observation supports Moreno’s (2006) cognitive-affective theory of learning with media, which posits that motivational factors can mediate learning processes by increasing or decreasing cognitive engagement. The observation also reflects Plass and Kaplan’s (2016) integrated cognitive-affective model of learning with multimedia, maintaining that affective processes are inseparable from cognitive processes. Empirical research has also shown that individuals can recognise the emotions portrayed by an instructor in video lectures (Lawson et al., 2021b), and they perform better on learning outcome tests when the instructor in the video is more emotionally appealing (Lawson et al., 2021a).

Several reviewed studies attempted to identify the affective processes during video-based learning. Specifically, social presence (Colliot & Jamet, 2018), situational interest (Zhang et al., 2020), and emotional engagement data (Wang et al., 2020b) were assessed alongside eye tracking metrics to identify aspects of affective experiences that were not captured by eye tracking devices. Survey questionnaires and Likert scales were employed to ask respondents to self-report their affective state after watching a video. Although self-report measures have advantages such as low cost, accessibility, and ease of administration, they have limitations when evaluating affective processes in video-based learning. In a multimedia learning context, individuals’ affective experiences tend to be continuous and dynamic. The affective state captured by filling out the survey at a single point in time (see for example Deng et al., 2020) does not represent learners’ emotional experiences in the entire learning process, nor can it reflect the ups and downs of the affective process (Plass et al., 2014). A meta-analysis failed to identify the mechanism by which colours and anthropomorphisms affect learning outcomes, because most multimedia learning studies did not use continuous process measures or conduct mediation analyses (Brom et al., 2018).

Moreover, affective processes differ from cognitive ones. Learners are capable of reconstructing cognitive processes, whereas the temporality of emotions suggests that the reconstruction of affective processes is more difficult (Le et al., 2018).

Furthermore, learners can be influenced by social desirability bias, speculate on research objectives, and hide their true feelings, further affecting the accuracy of self-reported ratings of experienced emotions. In view of these potential limitations, future research could employ physiological measurements, such as facial electromyography (Lackmann et al., 2021) and electrocardiography (Parong & Mayer, 2021), to detect individuals' affective experiences and combine cognitive, emotional, and behavioural data to further identify factors that facilitate or impede video-based learning.

Third, this review highlights the necessity of improving ecological validity when conducting eye tracking research on video-based learning. The results showed that eight studies required learners to sit in front of an eye tracking device with their head positioned in a chin rest to minimise head movement (e.g. van Marlen et al., 2018), and eleven studies did not allow learners to pause, rewind, or take notes while viewing a video (e.g. Wang et al., 2020b). Such research settings are very different from authentic learning contexts. Students confront various distractors and engage with other tasks while watching instructional videos in authentic learning environments (Alemdag, 2022). In autonomous, self-directed, and independent learning environments, such as flipped classrooms and MOOCs, it is extremely rare for learners to keep their heads still and be prohibited from interacting with the video content in any form. Interactivity is a key factor in the management of cognitive resources in the context of video-based learning, and the absence of interactivity in experimental research implies that recommendations for the design of instructional videos may not be generalised (Bétrancourt & Benetos, 2018).

This review also showed that the outcomes of video-based learning were assessed at various cognitive levels, such as knowledge retention, comprehension, and transfer. However, the reviewed studies did not disclose the learning objectives that the participants were expected to achieve after watching a video. This can deviate from natural teaching environments, where the intended outcomes for a lesson or unit are often stated before learning takes place (Biggs, 2014). To enhance ecological validity, it is contended that activities, time, physical space, roles, and perceptions be considered when designing and conducting research (Frey, 2018). Future research should strive to strengthen the degree of correspondence between the research settings and the phenomenon being investigated, such as using eye tracking systems that have a high tolerance for head movements, allowing learners to take control of the pacing of the video, and communicating learning objectives to participants at the beginning of the video.

Finally, this review highlights that boundary conditions must be considered when interpreting research findings. A revised 3P model was developed to reflect the key variables and relationships extracted from 44 eye tracking studies on video-based learning (Fig. 3). While scholars frequently manipulated the instructor's presence and social cues, non-social cues, and motivational strategies to predict learning-focused activities and learning outcomes, they tended to only provide descriptive reports of many student factors (e.g. age and occupation) and teaching context variables (e.g. topic and length of videos); to date, the correlations between these variables and other teaching and learning factors remain largely unknown. This is not consistent with the original 3P model and Biggs' proposition that teaching and

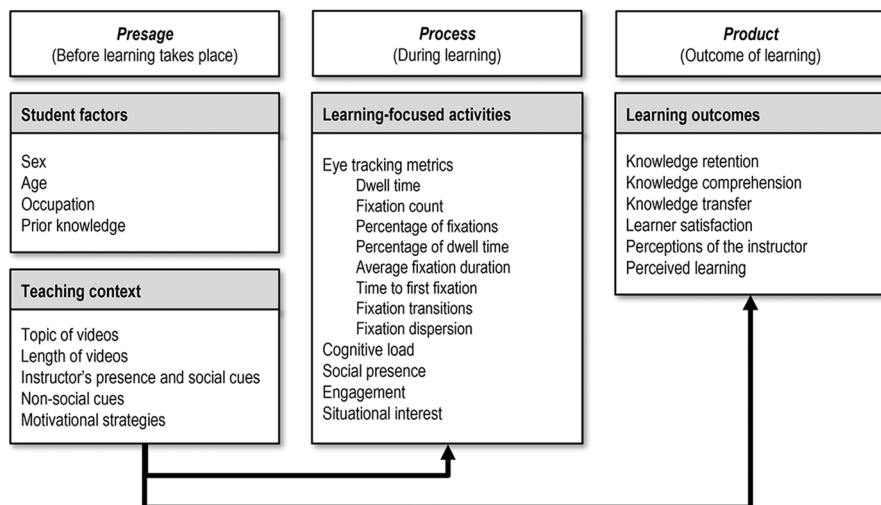


Fig. 3 Key variables and relationships extracted from the reviewed studies

learning factors are not static but interact with each other at different stages of learning (Biggs et al., 2001). The exploration of boundary conditions is intertwined with the theory development process—boundary conditions should not only be perceived as an amendment to theory but also a means for theory development (Busse et al., 2017). Some of the descriptively reported variables may be important boundary conditions in determining how video design principles work across different types of learners and teaching contexts.

The boundary conditions that require consideration include learner characteristics (Mayer, 2021), such as age and prior knowledge. This review provides unambiguous evidence that eye tracking research on video-based learning is primarily conducted with university students. This finding coincides with Alemdag and Cagiltay's (2018) and Çeken and Taşkın's (2022) observation that college students are the main participant group in broader multimedia learning research. Due to differences in age, literacy, and academic progress, video design principles that are effective for university students may not be equally efficacious for other learner categories. The literature has revealed that multimedia learning strategies that are effective for university students, such as providing graphical representations and cued texts (McTigue, 2009), incorporating motion and signalling in PowerPoint presentations (Schrader & Rapp, 2016), using peers for video explanation and demonstration (Hoogerheide et al., 2016), and training learners to actively link verbal and pictorial information (Hoch et al., 2021) have a limited effect on K-12 students. These differences highlight the importance of identifying boundary conditions and testing the video design principles that worked for university students in pre-college populations instead of indiscriminately applying pre-existing ones.

Other boundary conditions that require consideration are environmental factors, such as the topic of videos and type of knowledge. The topics of videos varied substantially across the reviewed studies, yet no studies have probed the interactive

effects of design principles with the topic of videos on student learning. This raises concerns about the transferability of video design principles. The preferred way of disseminating knowledge and skills through instructional videos can vary significantly across academic disciplines: arts and humanities disciplines show a predilection for a person-centric video style, science and technology disciplines favour a media-centric video style, and social sciences disciplines display a preference for a balance between person-centric and media-centric video styles (Santos Espino et al., 2016). These preferences are likely to be rooted in the intrinsic properties of disciplinary content. Future research could explore whether a design principle proven robust in a discipline holds valid in an entirely different area. A shift in the type of knowledge transmitted through instructional videos can also render an established design principle obsolete. For example, empirical research has shown that the instructor's image facilitates the learning of declarative knowledge but interferes with procedural knowledge (Hong et al., 2018). A recent meta-analysis has also shown that visual cues were most beneficial to learning performance when non-procedural tasks were taught in instructional videos (Xie et al., 2021). Practitioners need this contextual information to determine the boundary conditions for video design principles to be effective. Therefore, it is important for scholars to not only state for whom the video design principles are potentially effective but also explore the teaching circumstances and learning opportunities where these principles work the best.

5 Conclusion

Video-based learning is becoming increasingly prevalent in an era in which educational technology, multimedia learning, and the democratisation of education are valued in society (Madariaga et al., 2021; Sablić et al., 2021). To explain the underlying mechanisms that facilitate video-based learning, a growing number of empirical studies have used eye tracking technology to capture learners' unconscious, moment-to-moment processes when watching instructional videos. This begs consideration of how the utilisation of eye tracking technology and the interpretation of eye movement indicators have improved our understanding of the mechanisms underlying effective video-based learning and what caution needs to be exercised when interpreting findings of eye tracking research on video-based learning. To address these research questions, this review evaluated 44 eye tracking studies on video-based learning conducted between 2010 and 2021. The findings of this review suggest that eye tracking metrics are not a panacea for analysing non-conscious processes during video-based learning. Not all the reviewed studies successfully explained the mechanisms underlying effective video-based learning through utilising eye tracking technology, and few reviewed studies disentangled the complex relationship between eye tracking metrics and cognitive activities these metrics represent. While emotion plays a critical role in multimedia learning environments, learners' emotional processes and affective gains were often ignored. When interpreting findings of eye tracking research on video-based learning, it is imperative that researchers and practitioners pay close attention to the ecological validity and

boundary conditions of video design interventions that have been proven effective in existing studies.

This review has important implications for researchers. Specifically, its key findings inspire a number of propositions for designing and interpreting eye tracking research on video-based learning. Firstly, researchers should use additional psychophysiological and/or self-report measures in conjunction with eye tracking devices to explain the mechanisms underlying effective video-based learning. Similarly, it is critical that researchers use triangulation to disentangle the relationship between eye tracking metrics and the cognitive activities these metrics represent, rather than coming up with ad hoc explanations for the observed set of eye movement indicators. Secondly, researchers should consider capturing learners' continuous and dynamic affective processes while they watch instructional videos to investigate the emotional factors that may potentially mediate learning with instructional videos. Thirdly, it is important for researchers to strengthen the ecological validity of eye tracking research on video-based learning; for example, they may use eye tracking devices with a high tolerance for head movements, allow learners to take control of the pace of the video, and communicate the learning objectives of the video to participants. Finally, researchers should consider boundary conditions, including personal (e.g. learners' prior knowledge of videos) and environmental factors (e.g. the type of knowledge transmitted through videos) when interpreting findings. Practitioners such as teachers and video production managers also need this contextual information to determine when to apply established principles to the design and production of instructional videos.

6 Limitations

This review highlighted a number of opportunities for advancing the field. However, several limitations should be kept in mind when considering the findings of this review. While the 3P model of teaching and learning has provided a useful heuristic tool for organising and interrogating literature, adopting a different research framework may provide additional insights. In addition, this review adopted a deductive approach and superimposed the key variables in the model. For instance, given that situational interest is a psychological state characterised by increased affect, attention, and concentration during learner engagement (Quinlan, 2019) and is often operationalised as a learning process variable (Linnenbrink-Garcia et al., 2013), this review categorised situational interest as a process factor. This categorisation process may introduce ontological constraints. Other strategies may be used to explore this topic: future research could apply alternative frameworks or adopt a more inductive approach towards organising and analysing the literature. In addition, the selection criteria used in this review allowed the researchers to capture a representative selection of scientific studies on the topic of interest. An analysis of conference papers, dissertations, and sources published in languages other than English may yield slightly different results. Future studies should adopt different selection criteria to evaluate eye tracking research on video-based learning.

Authors' contributions RD collected the data, performed the analysis, and wrote the paper. YG validated the findings and revised the paper. All authors read and approved the final manuscript.

Funding This study is supported by the National Natural Science Foundation of China (Grant No. 72204072).

Data availability The datasets analysed during the current study are available from the first author on reasonable request.

Declarations

Human and animal rights and informed consent The study does not include any human participants or animals.

Conflict of interest The authors declare that they have no competing interests.

References

- Alemdag, E., (2022). Effects of instructor-present videos on learning, cognitive load, motivation, and social presence: A meta-analysis. *Education and Information Technologies*, 27, 12713–12742.
- Alemdag, E., & Cagiltay, K. (2018). A systematic review of eye tracking research on multimedia learning. *Computers & Education*, 125, 413–428.
- Altan, T., & Cagiltay, K. (2015). An eye-tracking analysis of spatial contiguity effect in educational animations. *International Conference on Learning and Collaboration Technologies, Los Angeles*.
- Anderson, J. R., Bothell, D., & Douglass, S. (2004). Eye movements do not reflect retrieval processes: limits of the eye-mind hypothesis. *Psychological Science*, 15(4), 225–231.
- Anmarkrud, Ø, Andresen, A., & Bråten, I. (2019). Cognitive load and working memory in multimedia learning: conceptual and measurement issues. *Educational Psychologist*, 54(2), 61–83.
- Arai, L., Britten, N., Popay, J., Roberts, H., Petticrew, M., Rodgers, M., & Sowden, A. (2007). Testing methodological developments in the conduct of narrative synthesis: a demonstration review of research on the implementation of smoke alarm interventions. *Evidence & Policy*, 3(3), 361–383.
- Arslan-Ari, I. (2018). Learning from instructional animations: how does prior knowledge mediate the effect of visual cues? *Journal of Computer Assisted Learning*, 34(2), 140–149.
- 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.
- Bender, L., Renkl, A., & Eitel, A. (2021). When and how seductive details harm learning. A study using cued retrospective reporting. *Applied Cognitive Psychology*, 35(4), 948–959.
- Benton, S. L., & Cashin, W. E. (2014). Student ratings of instruction in college and university courses. In M. B. Paulsen (Ed.), *Higher education: handbook of theory and research* (pp. 279–326). Springer.
- Bétrancourt, M., & Benetos, K. (2018). Why and when does instructional video facilitate learning? A commentary to the special issue “developments and trends in learning with instructional video”. *Computers in Human Behavior*, 89, 471–475.
- Biggs, J. B. (1993a). From theory to practice: a cognitive systems approach. *Higher Education Research & Development*, 12(1), 73–85.
- Biggs, J. B. (1993b). What do inventories of students' learning processes really measure? A theoretical review and clarification. *British Journal of Educational Psychology*, 63(1), 3–19.
- Biggs, J. B. (2003). *Teaching for quality learning at university* (2nd ed.). Open University Press.
- Biggs, J. B. (2014). Constructive alignment in university teaching. *HERDSA Review of Higher Education*, 1, 5–22.
- Biggs, J. B., Kember, D., & Leung, D. Y. P. (2001). The revised two-factor study process questionnaire: R-SPQ-2F. *British Journal of Educational Psychology*, 71(1), 133–149.
- Boucheix, J. M., Gauthier, P., Fontaine, J. B., & Jaffaux, S. (2018). Mixed camera viewpoints improve learning medical hand procedure from video in nurse training? *Computers in Human Behavior*, 89, 418–429.

- Boyle, E. A., MacArthur, E. W., Connolly, T. M., Hainey, T., Manea, M., Kärki, A., & van Rosmalen, P. (2014). A narrative literature review of games, animations and simulations to teach research methods and statistics. *Computers & Education*, 74, 1–14.
- Brom, C., Stárková, T., & D'Mello, S. K. (2018). How effective is emotional design? A meta-analysis on facial anthropomorphisms and pleasant colors during multimedia learning. *Educational Research Review*, 25, 100–119.
- Busse, C., Kach, A. P., & Wagner, S. M. (2017). Boundary conditions: what they are, how to explore them, why we need them, and when to consider them. *Organizational Research Methods*, 20(4), 574–609.
- Calvo-Ferrer, J. R. (2017). Educational games as stand-alone learning tools and their motivational effect on L2 vocabulary acquisition and perceived learning gains. *British Journal of Educational Technology*, 48(2), 264–278.
- Çeken, B., & Taşkın, N. (2022). Multimedia learning principles in different learning environments: a systematic review. *Smart Learning Environments*, 9(1), 1–22.
- Chen, C. M., & Wang, H. P. (2011). Using emotion recognition technology to assess the effects of different multimedia materials on learning emotion and performance. *Library & Information Science Research*, 33(3), 244–255.
- Chisari, L. B., Mockevičiūtė, A., Ruitenburt, S. K., Vemde, L., Kok, E. M., & Gog, T. (2020). Effects of prior knowledge and joint attention on learning from eye movement modelling examples. *Journal of Computer Assisted Learning*, 36(4), 569–579.
- Cojean, S., & Jamet, E. (2022). Does an interactive table of contents promote learning from videos? A study of consultation strategies and learning outcomes. *British Journal of Educational Technology*, 53, 269–285.
- 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.
- Conradie, J., & Frith, J. (2000). Comprehension tests in mathematics. *Educational Studies in Mathematics*, 42(3), 225–235.
- Cook, D. J., Mulrow, C. D., & Haynes, R. B. (1997). Systematic reviews: synthesis of best evidence for clinical decisions. *Annals of Internal Medicine*, 126(5), 376–380.
- de Koning, B. B., Tabbers, H. K., Rikers, R. M. J. P., & Paas, F. (2010). Attention guidance in learning from a complex animation: seeing is understanding? *Learning and Instruction*, 20(2), 111–122.
- Deng, R., & Benckendorff, P. (2021). What are the key themes associated with the positive learning experience in MOOCs? An empirical investigation of learners' ratings and reviews. *International Journal of Educational Technology in Higher Education*, 18, 9.
- Deng, R., Benckendorff, P., & Gannaway, D. (2019). Progress and new directions for teaching and learning in MOOCs. *Computers & Education*, 129, 48–60.
- Deng, R., Benckendorff, P., & Gannaway, D. (2020). Learner engagement in MOOCs: scale development and validation. *British Journal of Educational Technology*, 51(1), 245–262.
- Deng, R., Benckendorff, P., & Gao, Y. (2022). Limited usefulness of learning style instruments in advancing teaching and learning. *The International Journal of Management Education*, 20(3), 100686.
- Dunkin, M. J., & Biddle, B. J. (1974). *The study of teaching*. Holt, Rinehart & Winston.
- Expósito, A., Sánchez-Rivas, J., Gómez-Calero, M. P., & Pablo-Romero, M. P. (2020). Examining the use of instructional video clips for teaching macroeconomics. *Computers & Education*, 144, 103709.
- Faber, M., Krasich, K., Bixler, R. E., Brockmole, J. R., & D'Mello, S. K. (2020). The eye-mind wandering link: identifying gaze indices of mind wandering across tasks. *Journal of Experimental Psychology: Human Perception and Performance*, 46(10), 1201–1221.
- 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.
- Fiorella, L., van Gog, T., Hoogerheide, V., & Mayer, R. E. (2017). It's all a matter of perspective: viewing first-person video modeling examples promotes learning of an assembly task. *Journal of Educational Psychology*, 109(5), 653–665.
- Fredricks, J. A., & McColskey, W. (2012). The measurement of student engagement: a comparative analysis of various methods and student self-report instruments. In S. L. Christenson, A. L. Reschly, & C. Wylie (Eds.), *Handbook of research on student engagement* (pp. 319–339). Springer.
- Frey, B. (2018). *The SAGE encyclopedia of educational research, measurement, and evaluation*. SAGE.

- Garrett, N. (2021). Segmentation's failure to improve software video tutorials. *British Journal of Educational Technology*, 52(1), 318–336.
- Gegenfurtner, A., Lehtinen, E., Jarodzka, H., & Säljö, R. (2017). Effects of eye movement modeling examples on adaptive expertise in medical image diagnosis. *Computers & Education*, 113, 212–225.
- Geisen, E., & Romano Bergstrom, J. (2017). Developing the usability testing protocol. In E. Geisen, & J. Romano Bergstrom (Eds.), *Usability testing for survey research* (pp. 111–129). Morgan Kaufmann.
- Gordon, P. C., Hendrick, R., Johnson, M., & Lee, Y. (2006). Similarity-based interference during language comprehension: evidence from eye tracking during reading. *Journal of Experimental Psychology: Learning Memory and Cognition*, 32(6), 1304–1321.
- Harnish, R. J., & Bridges, K. R. (2011). Effect of syllabus tone: students' perceptions of instructor and course. *Social Psychology of Education*, 14(3), 319–330.
- Hoch, E., Scheiter, K., Stalbovs, K., & Gerjets, P. (2021). The intention was good: How promoting strategy use does not improve multimedia learning for secondary students. *British Journal of Educational Psychology*, 91, 1291–1309.
- 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.
- Hoogerheide, V., Loyens, S. M. M., & van Gog, T. (2015). Learning from video modeling examples: does gender matter? *Instructional Science*, 44(1), 69–86.
- Hoogerheide, V., van Wermeskerken, M., Loyens, S. M. M., & van Gog, T. (2016). Learning from video modeling examples: Content kept equal, adults are more effective models than peers. *Learning and Instruction*, 44, 22–30.
- Jamet, E. (2014). An eye-tracking study of cueing effects in multimedia learning. *Computers in Human Behavior*, 32, 47–53.
- Jang, D. M., Yang, I. H., & Kim, S. U. (2020). Detecting mind-wandering based on eye-tracking data during a physics video lecture. *Education Sciences*, 10, 1–13.
- Jarodzka, H., Balslev, T., Holmqvist, K., Nyström, M., Scheiter, K., Gerjets, P., & Eika, B. (2012). Conveying clinical reasoning based on visual observation via eye-movement modelling examples. *Instructional Science*, 40(5), 813–827.
- Jarodzka, H., Scheiter, K., Gerjets, P., & van Gog, T. (2010). In the eyes of the beholder: how experts and novices interpret dynamic stimuli. *Learning and Instruction*, 20(2), 146–154.
- Jarodzka, H., van Gog, T., Dorr, M., Scheiter, K., & Gerjets, P. (2013). Learning to see: guiding students' attention via a Model's eye movements fosters learning. *Learning and Instruction*, 25, 62–70.
- Jensen, J. L., Holt, E. A., Sowards, J. B., Ogden, H., & West, R. E. (2018). Investigating strategies for pre-class content learning in a flipped classroom. *Journal of Science Education and Technology*, 27(6), 523–535.
- Just, M. A., & Carpenter, P. A. (1980). A theory of reading: from eye fixations to comprehension. *Psychological Review*, 87(4), 329–354.
- Kok, E. M., & Jarodzka, H. (2017). Before your very eyes: the value and limitations of eye tracking in medical education. *Medical Education*, 51(1), 114–122.
- 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.
- Krathwohl, D. R. (2002). A revision of Bloom's taxonomy: an overview. *Theory into Practice*, 41(4), 212–218.
- Krebs, M. C., Schüler, A., & Scheiter, K. (2019). Just follow my eyes: the influence of model-observer similarity on eye movement modeling examples. *Learning and Instruction*, 61, 126–137.
- Kruger, J. L., & Steyn, F. (2014). Subtitles and eye tracking: Reading and performance. *Reading Research Quarterly*, 49(1), 105–120.
- Lackmann, S., Léger, P. M., Charland, P., Aubé, C., & Talbot, J. (2021). The influence of video format on engagement and performance in online learning. *Brain Sciences*, 11(2), 128.
- Lawson, A. P., Mayer, R. E., Adamo-Villani, N., Benes, B., Lei, X., & Cheng, J. (2021a). The positivity principle: do positive instructors improve learning from video lectures? *Educational Technology Research and Development*, 69(6), 3101–3129.
- Lawson, A. P., Mayer, R. E., Adamo-Villani, N., Benes, B., Lei, X., & Cheng, J. (2021b). Recognizing the emotional state of human and virtual instructors. *Computers in Human Behavior*, 114, 106554.
- Le, Y., Liu, J., Deng, C., & Dai, D. Y. (2018). Heart rate variability reflects the effects of emotional design principle on mental effort in multimedia learning. *Computers in Human Behavior*, 89, 40–47.

- Linnenbrink-Garcia, L., Patall, E. A., & Messersmith, E. E. (2013). Antecedents and consequences of situational interest. *British Journal of Educational Psychology*, 83, 591–614.
- Madariaga, L., Nussbaum, M., Gutiérrez, I., Barahona, C., & Meneses, A. (2021). Assessment of user experience in video-based learning environments: From design guidelines to final product. *Computers & Education*, 167, 104176.
- Madsen, J., Júlio, S. U., Gucik, P. J., Steinberg, R., & Parra, L. C. (2021). Synchronized eye movements predict test scores in online video education. *PNAS*, 118(5), Article e2016980118.
- Mayer, R. E. (2021). Evidence-based principles for how to design effective instructional videos. *Journal of Applied Research in Memory and Cognition*, 10(2), 229–240.
- McTigue, E. M. (2009). Does multimedia learning theory extend to middle-school students? *Contemporary Educational Psychology*, 34(2), 143–153.
- Montero Perez, M., Peters, E., & Desmet, P. (2015). Enhancing vocabulary learning through captioned video: an eye-tracking study. *The Modern Language Journal*, 99(2), 308–328.
- Moreno, R. (2006). Does the modality principle hold for different media? A test of the method-affects-learning hypothesis. *Journal of Computer Assisted Learning*, 22(3), 149–158.
- Neta, M., Tong, T. T., Rosen, M. L., Enersen, A., Kim, M. J., & Dodd, M. D. (2017). All in the first glance: first fixation predicts individual differences in valence bias. *Cognition and Emotion*, 31(4), 772–780.
- Ouwehand, K., Gog, T., & Paas, F. (2015). Designing effective video-based modeling examples using gaze and gesture cues. *Educational Technology & Society*, 18, 78–88.
- Pan, S. C., Sana, F., Schmitt, A. G., & Bjork, E. L. (2020). Pretesting reduces mind wandering and enhances learning during online lectures. *Journal of Applied Research in Memory and Cognition*, 9(4), 542–554.
- Parong, J., & Mayer, R. E. (2021). Learning about history in immersive virtual reality: does immersion facilitate learning? *Educational Technology Research and Development*, 69(3), 1433–1451.
- Phillips, N. E., Ralph, B. C. W., Carriere, J. S. A., & Smilek, D. (2016). Examining the influence of saliency of peer-induced distractions on direction of gaze and lecture recall. *Computers & Education*, 99, 81–93.
- Pi, Z., Chen, M., Zhu, F., Yang, J., & Hu, W. (2022). Modulation of instructor's eye gaze by facial expression in video lectures. *Innovations in Education and Teaching International*, 59, 15–23.
- Pi, Z., Tang, M., & Yang, J. (2020a). Seeing others' messages on the screen during video lectures hinders transfer of learning. *Interactive Learning Environments*, 1–14.
- Pi, Z., Xu, K., Liu, C., & Yang, J. (2020b). Instructor presence in video lectures: Eye gaze matters, but not body orientation. *Computers & Education*, 144, 103713.
- 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.
- 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(4), 1020–1029.
- Pi, Z., Zhang, Y., Yang, J., Hu, W., & Yang, H. H. (2019). All roads lead to Rome: instructors' pointing and depictive gestures in video lectures promote learning through different patterns of attention allocation. *Journal of Nonverbal Behavior*, 43(4), 549–559.
- Pi, Z., Zhang, Y., Zhu, F., Chen, L., Guo, X., & Yang, J. (2021). The mutual influence of an instructor's eye gaze and facial expression in video lectures. *Interactive Learning Environments*, 1–18.
- Plass, J. L., Heidig, S., Hayward, E. O., Homer, B. D., & Um, E. (2014). Emotional design in multimedia learning: Effects of shape and color on affect and learning. *Learning and Instruction*, 29, 128–140.
- Plass, J. L., & Kaplan, U. (2016). Emotional design in digital media for learning. In S. Tettegah, & M. Gartmeier (Eds.), *Emotions, Technology, Design, and Learning* (pp. 131–162). Academic.
- Popay, J., Roberts, H., Sowden, A., Petticrew, M., Britten, N., Arai, L., Roen, K., & Rodgers, M. (2005). Developing guidance on the conduct of narrative synthesis in systematic reviews. *Journal of Epidemiology and Community Health*, 59, A7.
- Quinlan, K. M. (2019). What triggers students' interest during higher education lectures? Personal and situational variables associated with situational interest. *Studies in Higher Education*, 44(10), 1781–1792.
- Rogaten, J., Rienties, B., Sharpe, R., Cross, S., Whitelock, D., Lygo-Baker, S., & Littlejohn, A. (2019). Reviewing affective, behavioural and cognitive learning gains in higher education. *Assessment & Evaluation in Higher Education*, 44(3), 321–337.

- Sablić, M., Miroslavljević, A., & Škugor, A. (2021). Video-based learning (VBL)—Past, present and future: an overview of the research published from 2008 to 2019. *Technology Knowledge and Learning*, 26, 1061–1077.
- Santos Espino, J. M., Suárez, A., & Artal, G. (2016). Speakers and boards: a survey of instructional video styles in MOOCs. *Technical Communication*, 63(2), 101–115.
- Schindler, M., & Lilienthal, A. J. (2019). Domain-specific interpretation of eye tracking data: towards a refined use of the eye-mind hypothesis for the field of geometry. *Educational Studies in Mathematics*, 101(1), 123–139.
- 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.
- Schrader, P. G., & Rapp, E. E. (2016). Does multimedia theory apply to all students? The impact of multimedia presentations on science learning. *Journal of Learning and Teaching in Digital Age*, 1(1), 32–46.
- Schraw, G., Flowerday, T., & Lehman, S. (2001). Increasing situational interest in the classroom. *Educational Psychology Review*, 13(3), 211–224.
- Seel, N. M. (2012). Retention test. In N. M. Seel (Ed.), *Encyclopedia of the Sciences of Learning* (pp. 2859–2859). Springer.
- Sharafi, Z., Soh, Z., & Guéhéneuc, Y. G. (2015). A systematic literature review on the usage of eye-tracking in software engineering. *Information and Software Technology*, 67, 79–107.
- Short, J., Williams, E., & Christie, B. (1976). *The social psychology of telecommunications*. Wiley.
- 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.
- Sweller, J., van Merriënboer, J. J. G., & Paas, F. G. W. C. (1998). Cognitive architecture and instructional design. *Educational Psychology Review*, 10(3), 251–296.
- Tarchi, C., Zaccoletti, S., & Mason, L. (2021). Learning from text, video, or subtitles: a comparative analysis. *Computers & Education*, 160, 104034.
- Tullis, T., & Albert, B. (2013). *Measuring the user experience*. Morgan Kaufmann.
- van Gog, T., Jarodzka, H., Scheiter, K., Gerjets, P., & Paas, F. (2009). Attention guidance during example study via the model's eye movements. *Computers in Human Behavior*, 25(3), 785–791.
- 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.
- van Marlen, T., van Wermeskerken, M., Jarodzka, H., & van Gog, T. (2016). Showing a model's eye movements in examples does not improve learning of problem-solving tasks. *Computers in Human Behavior*, 65, 448–459.
- van Marlen, T., van Wermeskerken, M., Jarodzka, H., & van Gog, T. (2018). Effectiveness of eye movement modeling examples in problem solving: the role of verbal ambiguity and prior knowledge. *Learning and Instruction*, 58, 274–283.
- 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.
- van Wermeskerken, M., & van Gog, T. (2017). Seeing the instructor's face and gaze in demonstration video examples affects attention allocation but not learning. *Computers & Education*, 113, 98–107.
- Wang, C. Y., Tsai, M. J., & Tsai, C. C. (2016). Multimedia recipe reading: Predicting learning outcomes and diagnosing cooking interest using eye-tracking measures. *Computers in Human Behavior*, 62, 9–18.
- Wang, F., Li, W., Mayer, R. E., & Liu, H. (2018). Animated pedagogical agents as aids in multimedia learning: Effects on eye-fixations during learning and learning outcomes. *Journal of Educational Psychology*, 110(2), 250–268.
- Wang, F., Zhao, T., Mayer, R. E., & Wang, Y. (2020a). Guiding the learner's cognitive processing of a narrated animation. *Learning and Instruction*, 69, 1–12.
- Wang, J., Antonenko, P., & Dawson, K. (2020b). Does visual attention to the instructor in online video affect learning and learner perceptions? An eye-tracking analysis. *Computers & Education*, 146, 1–16.
- Wang, J., Antonenko, P., Keil, A., & Dawson, K. (2020c). Converging subjective and psychophysiological measures of cognitive load to study the effects of instructor-present video. *Mind Brain and Education*, 14(3), 279–291.

- Wang, X., Lin, L., Han, M., & Spector, J. M. (2020d). Impacts of cues on learning: using eye-tracking technologies to examine the functions and designs of added cues in short instructional videos. *Computers in Human Behavior*, 107, 106279.
- Wang, H., Pi, Z., & Hu, W. (2019a). The instructor's gaze guidance in video lectures improves learning. *Journal of Computer Assisted Learning*, 35(1), 42–50.
- Wang, J., Antonenko, P., Celepkolu, M., Jimenez, Y., Fieldman, E., & Fieldman, A. (2019b). Exploring relationships between eye tracking and traditional usability testing data. *International Journal of Human–Computer Interaction*, 35(6), 483–494.
- 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.
- Wong, R. M., & Adesope, O. O. (2021). Meta-analysis of emotional designs in multimedia learning: a replication and extension study. *Educational Psychology Review*, 33, 357–385.
- Wu, C. J., & Liu, C. Y. (2022). Refined use of the eye-mind hypothesis for scientific argumentation using multiple representations. *Instructional Science*, 50, 551–569.
- Xiang, X. M., & Miller, D. (2020). Effects of simulation video on parental recall of seizure first aid: a quality improvement project. *Journal of Child Neurology*, 35(13), 908–911.
- Xie, H., Zhao, T., Deng, S., Peng, J., Wang, F., & Zhou, Z. (2021). Using eye movement modelling examples to guide visual attention and foster cognitive performance: a meta-analysis. *Journal of Computer Assisted Learning*, 37(4), 1194–1206.
- 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.
- 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.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.