Avatars vs. Video Presence: Effects of Instructor Presence on Cognitive Load in Video-Based Learning

Yuli Sutoto Nugroho[1,3], Marie-Luce Bourguet[1], Hamit Soyel[1] and Isabelle Mareschal[2]
[1]School of Electronic Engineering and Computer Science, Queen Mary University of London, UK
[2]School of Biological and Behavioural Sciences, Queen Mary University of London, UK
[3]Faculty of Engineering, Universitas Negeri Surabaya, Indonesia

y.nugroho@qmul.ac.uk (corresponding author)
marie-luce.bourguet@qmul.ac.uk
h.soyel@qmul.ac.uk
i.mareschal@qmul.ac.uk

https://doi.org/10.34190/ejel.23.3.3964

Abstract: This study examines the impact of instructor presence on cognitive load and learning outcomes in video-based learning environments, addressing a significant gap in optimising instructional design for digital education. Utilising eye-tracking technology to measure pupil dilation, a reliable proxy for cognitive load, this research compares three experimental conditions: no instructor presence, physical instructor presence, and avatar-based instructor presence. Thirty-three undergraduate participants engaged with educational videos, and their cognitive load was assessed through pupil dilation while learning gains were evaluated using pre- and post-tests. Findings indicate that physical instructor presence induces the highest extraneous cognitive load due to non-verbal distractions, negatively affecting learning outcomes. Conversely, avatar-based instructor presence effectively balances cognitive demands by reducing extraneous load and fostering germane cognitive processing, enhancing learning outcomes. The absence of an instructor minimises distractions and moderates cognitive load but results in only moderate learning gains, highlighting the importance of instructor presence in video-based learning, particularly for complex materials requiring contextual support and guided instruction. This research underscores the potential of avatars as a scalable and efficient instructional tool, especially in remote and asynchronous learning contexts. By simplifying visual cues and employing purposeful gestures, avatars mitigate extraneous distractions while maintaining instructional presence. These findings suggest that avatars can bridge the gap between the absence of an instructor and the potential overload associated with physical instructors. The study also demonstrates the value of integrating physiological measures like eye tracking into educational research to refine instructional designs further. This approach offers real-time insights into cognitive processing and learner responses, reducing biases inherent in self-reported measures. This work contributes actionable insights into designing scalable, effective educational technologies that optimise cognitive load and improve learning outcomes, paving the way for innovative approaches in modern e-learning.

Keywords: Cognitive load, Video-Based learning, Eye-Tracking, Avatar, Educational technology

1. Introduction

Effective learning materials, delivery methods, and environments are crucial in helping students meet their educational goals. Learning materials can be presented in various formats, with video emerging as a particularly impactful medium. Video-Based Learning (VBL) allows learners to control their pace by pausing, rewinding, or replaying content. By integrating audiovisual content, videos have proven to be an excellent tool for improving understanding and retention of information, surpassing traditional methods like reading or listening to lectures. (Sablić et al., 2021) (Gómez-Ortega et al., 2023). Examples include lecture videos in online courses and tutorial videos on platforms like YouTube or Coursera.

Massive Open Online Courses (MOOCs) have accelerated VBL adoption by providing scalable, interactive, and globally accessible education (Bettiol, Psereckis & MacIntyre 2022). These platforms make abstract concepts more tangible through animations, demonstrations, and real-world case studies. VBL has recently seen an increase in adoption, with many students participating in online courses.

Several studies have shown that videos, when designed according to multimedia learning principles, enhance retention and understanding more effectively than reading alone (Tarchi, Zaccoletti & Mason 2021; Alhazmi 2024). Compared to text or audio-only lectures, video-based learning is more engaging and compelling, promoting deeper learning by leveraging multisensory inputs to enhance cognitive processes (Lackmann et al. 2021). This aligns with Mayer's Cognitive Theory of Multimedia Learning, highlighting how well-designed videos reduce cognitive load and improve comprehension (Mayer 2009).

Cite this article: Nugroho, Y.S. et al. 2025. "Avatars vs. Video Presence: Effects of Instructor Presence on Cognitive Load in Video-Based Learning", *Electronic Journal of e-Learning*, 23(3), pp 79-91, https://doi.org/10.34190/ejel.23.3.3964

Cognitive load refers to the demand for working memory resources, which decreases as processing capacity increases (Lang 2006; Chen & Epps 2014). Cognitive Load Theory (CLT) guides effective teaching methods by addressing learning challenges through instructional design (Sweller & Chandler 1994). Cognitive Load (CL) consists of three domains: intrinsic CL, extraneous CL, and germane CL (Brünken, Seufert & Paas 2010). Intrinsic CL (Sweller & Chandler 1994) refers to the innate complexity of information and content that must be comprehended and learned. Extraneous CL is created by additional requirements that result from poor instructional design and are not directly related to the task (Sweller, Van Merrienboer & Paas 1998). Germane CL entails storing new information in long-term memory, allowing individuals to focus more on activity execution (Zhang et al. 2020). CLT highlights the need for instructional strategies that minimise extraneous load, optimise intrinsic load, and foster germane load to enhance learning, particularly in technology-driven education (Leppink et al. 2014). By aligning instructional design with cognitive capacity, CLT helps create more effective learning environments.

Some researchers use questionnaires to assess cognitive load, which can be subjective. More objective methods should be employed to evaluate the effects of various instructional designs on cognitive load. According to (Souchet et al. 2021), eye tracking is an effective method for measuring cognitive load, offering real-time, high-validity assessments. However, despite the growth in physiological signals measurement tools such as electroencephalography, eye tracking, and heart rate variability, research on their application in education is still sparse, with few studies exploring eye-tracking in learning contexts (Lim, Mountstephens & Teo 2020). This technology records eye positions, gaze patterns, fixations (where the eyes stop to focus), saccades (quick eye movement between points), and pupil dilation (how wide the pupil opens), which is particularly useful for evaluating cognitive load (Zagermann, Pfeil & Reiterer 2016; Bourguet et al. 2020).

(Negi & Mitra 2020) found that fixation duration can be an indicator for measuring cognition and attention. Nevertheless, fixation duration does not always indicate cognitive load, as other factors, such as visual salience, can also cause participants to fixate on a particular area. In this study, we focus on using pupil dilation as the indicator of cognitive load. Pupil dilation correlates with cognitive effort: a larger dilation indicates a higher cognitive load, while a smaller dilation suggests a lower strain (Peysakhovich, Dehais & Causse 2015; Zheng et al. 2022). The sympathetic nervous system controls pupil dilation, typically during mentally demanding or emotional activities. Baseline pupil size can also influence cognitive performance, as larger baselines are linked to reduced task-related dilation (Gilzenrat et al. 2010). Fixation duration, measured in milliseconds, indicates time spent viewing an “Area of Interest”, helping researchers understand cognitive load distribution (Zagermann et al. 2016). Areas of Interest (AOIs) help analyse visual attention by mapping eye movements to specific stimuli (Hessels et al. 2018). The accuracy of AOI definitions depends on eye-tracking precision. In this study, AOIs are used to examine how participants allocate attention in instructional videos, assessing the correlation between fixation transitions and pupil size increases. Incorporating such tools can provide insights into learner responses and inform the use of different learning delivery methods in distance education, especially VBL.

Instructional design plays a crucial role in VBL effectiveness. Optimising VBL implementation is essential to maximise learning outcomes. For example, the effects of different video lecture presentation formats on learning outcomes have not been thoroughly examined. Research on how "presence" in video content affects cognitive load remains limited. While prior studies have explored the role of physical instructors in increasing cognitive load (Polat 2023), few have systematically compared this with avatar-based representations and their potential to balance intrinsic and extraneous cognitive loads in digital environments (Pignatiello et al. 2019).

"Presence" in educational videos refers to an instructor’s perceived involvement in the learning material(Beege et al. 2023). The different types of presence in VBL are: physical instructor presence, where the real lecturer appears in the video with natural gestures and expressions; avatar presence, a simplified computer-generated representation of the lecturer using minimal, purposeful gestures; and no lecturer presence, which features only slides and voice-over without any visual representation of the instructor. Each mode influences cognitive load and learning effectiveness differently. A physical instructor can enhance motivation, but may increase cognitive load due to additional stimuli such as spontaneous gestures (Heidig et al. 2024). In contrast, avatars offer simplified, controlled visuals and balance cognitive demands more effectively (Ayres & Paas 2007). The absence of a lecturer removes social cues, allowing full attention to content, but might hinder learning motivation (Chi 2023). Gestures serve as non-verbal cues that clarify concepts, direct attention, and reinforce learning (Wakefield et al. 2018). Effective gestures reduce cognitive load by aiding comprehension, while excessive or irrelevant gestures can distract learners (Rohrer, Delais-Roussarie & Prieto 2020). In VBL, simplified, purposeful gestures (whether from an instructor or an avatar) enhance comprehension without overwhelming learners (Dargue, Sweller & Jones 2019).

Our research investigates how different types of instructor presence in videos, such as non-lecturer presence, physical presence, or presence as an avatar, affect learners' cognitive load and learning gain. Understanding these differences is key to designing effective VBL environments. Despite the increasing use of VBL, it remains unclear how different types of instructor presence affect learners' cognitive processing and learning outcomes. Previous studies have concentrated on a single form of instructor presence, highlighting a significant design gap, specifically, the limit of lecturer presence variation. Moreover, cultural differences in communication preferences may modulate how learners perceive and process instructional presence, potentially influencing cognitive load and learning outcomes. Addressing these gaps, the current study seeks to broaden the scope of previous research by empirically examining multiple types of instructor presence. To achieve this, it employs physiological measures to assess cognitive and learning effects, providing actionable insights for improving VBL instructional design. Accordingly, this study investigates the interplay between instructor presence, cognitive load, and learning gain, where learning gain refers to improved learners' knowledge or understanding after studying activities, measured by comparing pre- and post-test scores. The learning gain reflects how much the learner has learned due to educational intervention.

Our study is guided by the following research questions: How does the instructor's presence influence the learner's cognitive load? (RQ1); and What is the relationship between cognitive load and learning gain in VBL? (RQ2). By advancing our understanding of how lecturer presence influences cognitive processing and learning outcomes, this research contributes to developing scalable and effective digital education solutions across diverse learning environments in higher education and professional training. The findings of this study are expected to offer practical insights, such as guiding instructional designers to adopt avatar-based instructors in video learning environments, which may help reduce extraneous cognitive load while maintaining instructional presence.

2. Methodology

2.1 Experimental Design

This study employed a 3×1 within-subject design, where all 33 participants experienced three conditions of instructor presence (physical instructor, avatar, and no lecturer), serving as their own control group. This design minimises individual differences by comparing each participant's performance across all conditions. The experiment design had the following conditions regarding lecturer presence as the independent variable.

- Non-lecturer presence in the video: The learning material is presented asynchronously (accompanied by the researcher) with voice-over slides, but the lecturer is not visible.
- Lecturer's physical presence in the video: The learning material is presented asynchronously, including the lecturer's physical presence, capturing their face and upper body.
- Lecturer's presence as an avatar in the video: This format is similar to the lecturer's physical presence, but the avatar represents the lecturer. The avatar appears simple and realistic with a static facial expression (no movement of eyes or mouth) and is dressed in formal clothing. The avatar performs basic gestures, such as raising a hand, pointing a finger, and looking downward, which are inspired by the physical presence of the lecturer. The avatar's appearance in the learning video is shown in Figure 1.

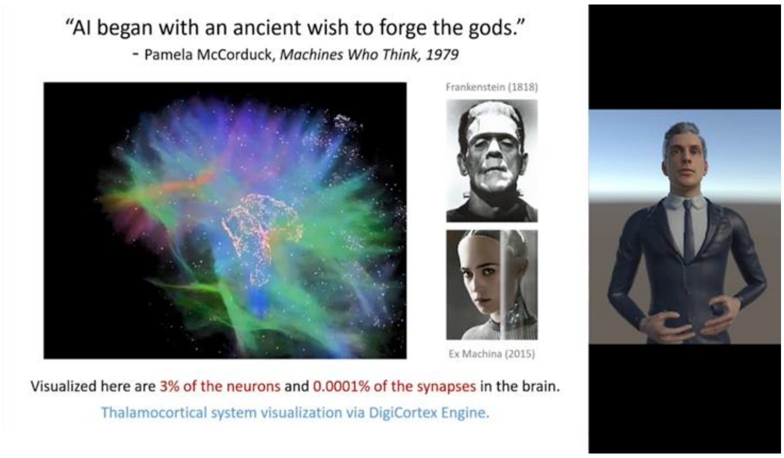

Figure 1: The appearance of the Avatar in the Video

The dependent variables were cognitive load, measured by the pupil diameter size increase or pupil dilation events and learning gain from pre- and post-test results. The total number of trials was 99 (=33 participants x 3 conditions).

The videos in this study were delivered asynchronously to avoid direct live interaction between participants and the instructor. This approach ensured that each participant received a consistent and controlled classroom experience. While synchronous settings might introduce richer social dynamics, they could also lead to inconsistencies in delivery, making it harder to isolate the effects of instructor presence on cognitive load.

2.2 Procedure and Video Stimuli

In this study, eye movements and pupil size were recorded from participants using eye-tracking technology under controlled conditions, including consistent lighting, environment, and timing. This experimental design ensured that pupil diameter measurements could be compared reliably across the different experimental conditions. The experiment took place in a university laboratory.

Three learning video stimuli were used, each lasting approximately 5 minutes. The videos were adapted from Lex Fridman's YouTube video on "Deep Learning" (www.youtube.com/watch?v=O5xeyoRL95U), chosen for its relevance to the participant's field of study. Additionally, the video layout, with the lecturer separate from the slides, allowed for seamless modification, enabling the controlled replacement of the instructor with an avatar and the deletion of the real lecturer. Overall, each participant spent about 60-75 minutes on the experiment. The visual timeline of the process is illustrated in Figure 2.

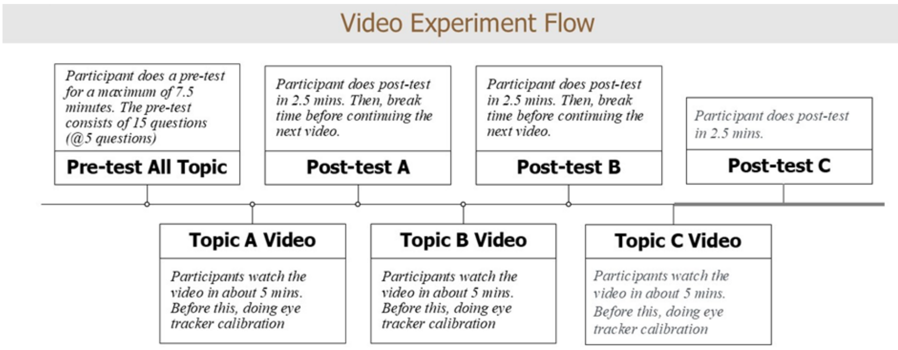

Figure 2: The Flow of Video Experiment

Before viewing the videos, participants completed a knowledge test of multiple-choice questions related to the content (pre-test). After watching the videos, they took a second knowledge test (post-test) to assess their learning gain.

The experiment began with 10 minutes of preparation and instruction, followed by a 7.5-minute pre-test (15 questions). Participants then underwent 5 minutes of eye-tracking calibration before watching a 5-minute video on Topic A ("Deep Learning Basics"), followed by a 2.5-minute post-test (same five questions from the pre-test).

After a 5-minute break, participants completed another 4-minute calibration and watched a 5-minute video on Topic B ("Representation Learning"), followed by another 2.5-minute post-test. A second 5-minute break preceded a final 4-minute calibration before watching a 5-minute video on Topic C ("The Challenge of Deep Learning"), concluding with a 2.5-minute post-test (same last five questions from the pre-test).

2.3 Participants

This study involved 33 participants. The sample size was determined based on a priori power analysis to ensure sufficient statistical power (80%) to detect medium effect sizes (Cohen's d = 0.5) with a significance level of 0.05 (Serdar et al. 2021). This is consistent with prior studies on cognitive load using eye-tracking technology in educational settings (Sáiz-Manzanares et al. 2024).

Participants were third-year undergraduate students majoring in Electronic Engineering and Computer Science, aged 20–23 years (M = 21.5, SD = 0.8). The same educational background guarantees that all participants have similar prior knowledge of the videos' topics. All individuals had either normal vision or vision corrected to normal, and no prior history of neurological disorders. Those specific criteria limited the number of participants. Nevertheless, the sample size was sufficient for the intended statistical analyses.

This experiment was conducted by the ethical guidelines. Detailed information for participants was provided in the advertisement email before they came to the study site. They had the right to withdraw at any time without explanation. Participants were asked to read and fill out the consent form at the beginning of the study.

Before the experiment, participants were briefed on the study's objectives and design in the laboratory. Study details were emailed beforehand, and participation was voluntary, allowing withdrawal at any time. No expenses were incurred by the participants in this research. Participants received £20-25, depending on their test scores, as compensation. This token of appreciation was given after the completion of the experiment.

2.4 Apparatus

Participants' eye movements and visual attention during video-based learning were recorded using eye-tracking technology. The collected data were analysed statistically to examine how different experimental conditions affect fixations and pupil dilation, which reflect visual attention and cognitive load. The eye-tracking device used was a 120 Hz core eyeglasses model from Pupil Labs (see Figure 3). Calibration was conducted before each session using a standard five-point calibration method to ensure accuracy. Data with a confidence level below 0.6 were excluded from the analysis.

Two laptops and an external monitor were used for the experiment. The primary laptop managed the setup, calibrated the eye-tracking system, and played video stimuli. A second laptop was assigned for participants to complete pre- and post-tests, ensuring independence from the investigator's control. The external monitor provided real-time eye-tracking data, allowing immediate accuracy verification and troubleshooting.

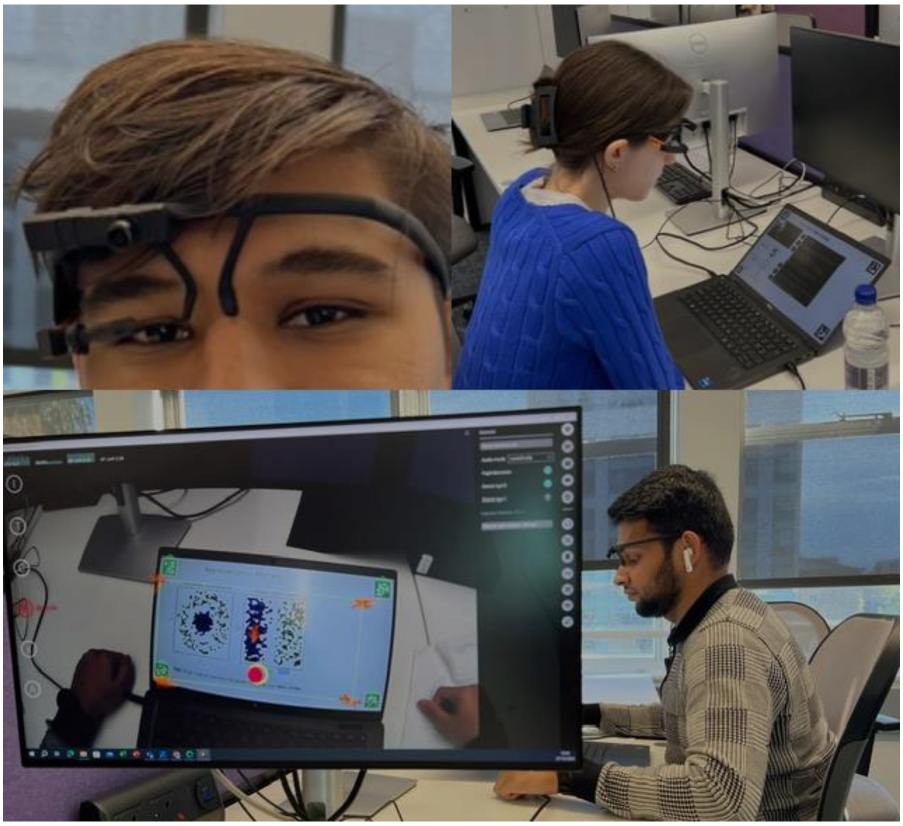

Figure 3: Experimental Tools (images used with authorisation from the study participants)

3. Results

3.1 Pupil Dilation Across Conditions

Pupil diameter was measured for each participant across the three experimental conditions. We assumed that an increase in pupil diameter, i.e., pupil dilation, corresponds to an increase in cognitive load. Subsequently, statistical analyses were performed to determine whether there were significant differences in pupil diameter changes across the three video stimuli.

Before processing the pupil diameter data, we first cleaned the raw data through several steps: (1) replacing blank diameter_3d values with the nearest non-blank value; (2) substituting data with low confidence (< 0.6) with the nearest high-confidence value; and (3) detecting outlier data using the z-score method and replacing them with the nearest non-outlier value.

Figure 4 illustrates the actual pupil diameter size. It plots the distribution of pupil diameter sizes for the 33 participants under the three experimental conditions: no lecturer presence, lecturer's physical presence, and lecturer present as an avatar. 14 people had a bigger pupil diameter with no lecturer presence, 14 people with a lecturer's physical presence, and 5 people with an avatar presence.

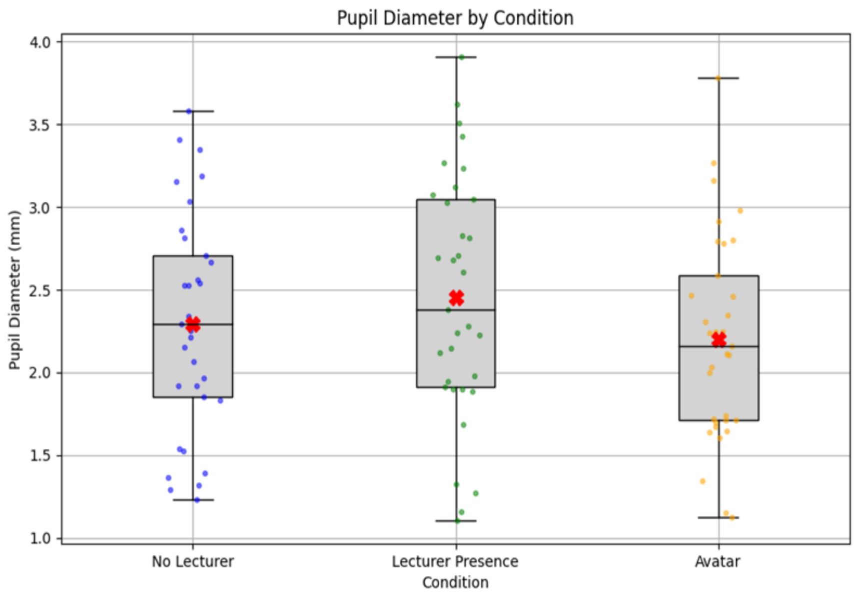

Figure 4: Pupil Diameter from all Participants across Conditions

Furthermore, to investigate the cognitive load experienced by participants across different experimental conditions, pupil dilation was measured, which was associated with the highest levels of cognitive load. To quantify pupil dilation, we first established a baseline by calculating the mean pupil diameter [1] and its variability through the standard deviation [2]. Significant pupil dilation was identified as any instance where the pupil diameter exceeded the baseline by more than one standard deviation [3], and the mean of significant dilation was computed [4]. A sample visualisation illustrating changes in pupil dilation over time for a single participant is presented in Figure 5.

$$\text{Baseline/Mean } (\bar{d}) = \frac{1}{N}\sum_{i=1}^{N} d_i \tag{1}$$

$$\text{Standard Deviation } (\sigma) = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(d_i - \bar{d})^2} \tag{2}$$

$$\text{Dilation Threshold} = \bar{d} + \sigma \tag{3}$$

$$\text{Significant Dilation (Mean)} = \frac{\sum_{i=1}^{N}[d_i > (\bar{d}+\sigma)]}{N_{\text{significant}}} \tag{4}$$

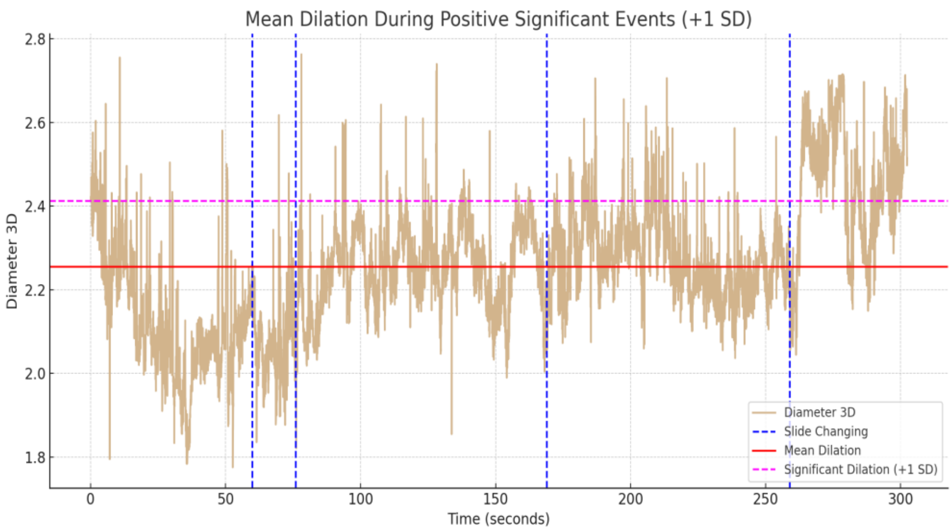

Figure 5: Example of Significant Pupil Dilation from One Participant

To see the statistical difference between lecturer conditions, we enriched the dataset by slicing it into slides on the videos. Using the within-subject design from 33 participants, topic A video had 4 slides, topic B had 6, and topic C had 5. Table 1 provides the slide design summary of each topic.

The mean significant dilation per slide for the non-lecturer presence is 2.31 mm; for the physical presence of the lecturer, it is 2.40 mm; and for the avatar lecturer, it is 2.01 mm. Initially, the Shapiro-Wilk test was performed to determine the normality of the data for each condition. The results indicated significant deviations from normality for all three conditions: no lecturer ($W = 0.893$, $p < 0.001$, $N = 165$), lecturer presence ($W = 0.895$, $p < 0.001$, $N = 165$), and an avatar lecturer ($W = 0.885$, $p < 0.001$, $N = 165$). These findings suggest that the data in all conditions are not normally distributed, justifying the use of non-parametric statistical methods for subsequent analyses.

Table 1: Topic Slide Design

Topic Video	Duration (min)	Number of Slides	
A	5	4	
B	5	6	
C	5	5	

Furthermore, we conducted the test comparing the significant pupil dilation between no lecturer and lecturer physical presence. We found that the p-value for a paired t-test indicates no significant difference between those conditions ($W = 5895.0$, $p = 0.687$, $N = 165$), with a small effect size ($r = -0.12$). Subsequently, we compared the conditions for no lecturer and the avatar. The test indicated a significant difference between those conditions ($W = 4983.0$, $p = 0.018$, $N = 165$), with a small to medium effect size ($r = -0.24$).

Finally, we examined the difference in significant pupil dilation between the physical presence of the lecturer and the avatar condition, which resulted ($W = 4756.0$, $p = 0.008$, $N = 165$), with a medium effect size ($r = -0.26$). To examine the effect of cognitive load in various forms of lecturer presence more closely, it is interesting to explore the correlation between pupil dilation and learning gain.

3.2 Correlation between Pupil Dilation and Learning Gain

Knowledge assessments were conducted before and after exposure to the instructional video stimuli to evaluate learning gains. These assessments included 15 multiple-choice questions to measure participants' understanding at each stage. Learning gains were calculated following the method described in (Marx & Cummings 2007). Specifically, when the post-test score was higher than the pre-test score, the difference was divided by the maximum possible score minus the pre-test score, as shown in (5). Conversely, if the post-test score was lower, the difference was divided by the pre-test score (6). Cases where the pre-test and post-test scores were identical and represented either the maximum or minimum score were excluded from the analysis(7). For instances where the pre-test and post-test scores matched but were not at extremes, the learning gain was set to 0 (8).

$\frac{post-pre}{Max.-pre}$, if $post > pre$ (5)

$\frac{post-pre}{pre}$, if $post < pre$ (6)

exclude, if $post = pre =$ maximum or minimum (7)

0, if $post = pre$ (8)

Table 2 presents the mean pupil dilation (in mm) and mean learning gain (in percentage) along with their respective standard deviations (SD) for each experimental condition. We first assessed the normality of significant pupil dilation data per condition and learning gain using the Shapiro-Wilk test to determine the appropriate correlation test.

Table 2: Mean Pupil Dilation and Learning Gains Across Conditions

Condition	Mean Pupil Dilation (mm)	SD	Mean Learning Gain (%)	SD
No Lecturer	2.3090	1.2142	45.6061	43.9080
Physical Lecturer	2.3986	1.4259	36.1616	44.7449
Avatar Lecturer	2.0073	1.3508	47.9293	39.2051

The assumption of normality for all variables was tested using the Shapiro-Wilk test. The results indicated that some variables deviated significantly from normality: learning gain of no lecturer presence ($W = 0.884, p = 0.002$), learning gain from the lecturer's physical presence ($W = 0.909, p = 0.009$), and learning gains from the avatar lecturer ($W = 0.921, p = 0.020$). Conversely, all pupil dilation variables were found to be normally distributed: pupil dilation of no lecturer ($W = 0.970, p = 0.488$), pupil dilation of the lecturer's physical presence ($W = 0.985, p = 0.921$), and pupil dilation of the avatar lecturer ($W = 0.966, p = 0.375$).

Given the significant deviations from normality in the learning gain variables, Spearman's rank-order correlation was employed to examine the relationship between learning gain and pupil dilation across conditions. The analysis yielded the following results:

- Learning Gain and Pupil Dilation in no lecturer presence: A small positive correlation was observed ($r_s = 0.132, p = 0.463$). This suggests that minimal instructional presence may require learners to engage their cognitive resources independently, though the weak correlation indicates a limited impact on learning outcomes.
- Learning Gain and Pupil Dilation in lecturer's physical presence: A moderate negative correlation was observed ($r_s = -0.335, p = 0.057$), which approached statistical significance. This finding suggests that the physical presence of a lecturer may introduce extraneous cognitive load through non-verbal cues and visual complexity.
- Learning Gain and Pupil Dilation in avatar lecturer: A small-to-moderate positive correlation was observed ($r_s = 0.224, p = 0.211$). This suggests that avatars help optimise germane cognitive load by providing sufficient instructional presence while minimising extraneous distractions, thus enhancing learning outcomes.

The findings suggest potential trends in the relationships between pupil dilation and learning gain. These results align with the study's hypotheses that different forms of instructor presence impact pupil dilation and learning gain differently. The correlations between learning gain and pupil dilation across the different conditions can be seen in Figure 6.

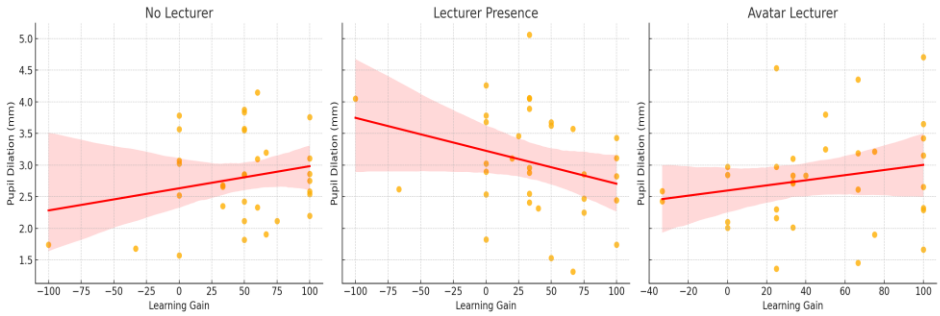

Figure 6: Correlation Between Pupil Dilation and Learning Gain from All Participants

3.3 Pupil Dilation Comparison Between High and Low Learning Gain Participant Groups

We categorised participants in each experimental condition into two groups based on their learning outcomes: high versus low learning gain. We compare the pupil dilation between high- and low-level learning gain in each condition to see if there is a relationship between high and low learning gain on pupil dilation in different presence conditions.

With 33 participants per condition, we ranked the participants based on the test's learning gain and time duration. We assigned the top 16 performers to the "high learning gain" group, eliminating 1 median participant, and the remaining 16 belonged to the "low learning gain" group. We enriched the data by measuring pupil dilation at intervals of 1 minute to strengthen the statistical calculation. Thus, each group has 80 data points (16 x 5).

None of the data points were normally distributed after performing a normality test on all the data points. Hence, the Wilcoxon signed-rank test was used to see the difference in pupil dilation size between high and low performers across three different presence conditions.

Firstly, we compared the pupil dilation between high and low learning gain groups in the no lecturer condition. The results indicated no significant difference in pupil dilation between those learning levels $W = 1527.0$, $p = 0.796$ ($N = 80$). This indicates that the absence of a lecturer results in a similar cognitive load across both high and low-learning gain groups, suggesting that learners rely on their intrinsic cognitive effort equally. Next, we analysed the lecturer's physical presence condition. We found a significant difference in pupil dilation between the high and low-learning gain groups $W = 947.0$, $p = 0.012$ ($N = 80$), indicating that excessive cognitive load may disproportionately hinder learners with lower initial performance levels. Interestingly, the low learning gain group exhibited a higher pupil dilation. Lastly, we examined pupil dilation when the lecturer was present as an avatar. The results revealed no significant difference $W = 1389.0$, $p = 0.702$ ($N = 80$). This suggests that the simplified and controlled representation of the avatar maintains a balanced cognitive load, benefiting both high and low-learning gain groups similarly.

3.4 Correlation Between Number of Transitions Between AOIs and Pupil Dilation

We explored the correlation between the number of fixation transitions and pupil diameter increase in the lecturer presence and avatar experimental conditions for all the participants (n = 33). Transition refers to switching visual attention from AOI 1 (slide) to AOI 2 (lecturer area); see Figure 1. This analysis aimed to determine whether frequent shifts in attention across different areas of stimuli contributed to pupil dilation. The scanpath shows an example of the transition from one participant, as seen in Figure 7.

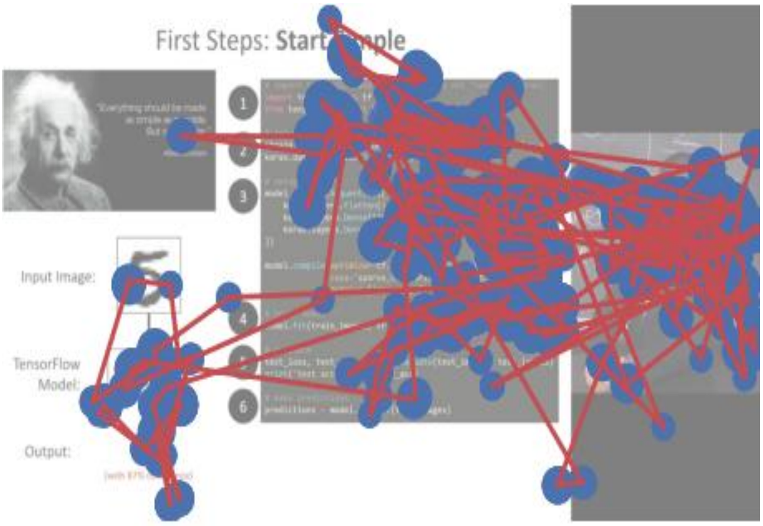

Figure 7: Example of Scanpath from One Participant

A Shapiro-Wilk test was conducted to evaluate normality. The results showed that the Transition in the lecturer's physical presence significantly deviates from normality ($W = 0.915, p = 0.014$), while Pupil Dilation of the lecturer's physical presence ($W = 0.985, p = 0.921$), Transition in the avatar lecturer ($W = 0.967, p = 0.401$), and Pupil Dilation of the avatar lecturer ($W = 0.966, p = 0.375$) followed normal distributions ($N = 33$ For all columns).

Based on these results, a Spearman correlation was used for Pupil Dilation and Transition in lecturer presence, yielding a small negative monotonic relationship ($r_s = -0.168, p = 0.351$). A Pearson correlation was used for Pupil Dilation and Transition in the avatar lecturer, indicating a small positive linear relationship ($r = 0.156, p = 0.385$).

4. Discussion

4.1 Effect of Instructor's Presence on Cognitive Load

The comparative analysis of pupil dilation distributions across the three conditions (no lecturer, lecturer presence, and avatar) suggests different levels of cognitive burden. The findings indicate that the physical presence of a lecturer is associated with the highest levels of cognitive load, as evidenced by the largest mean pupil dilation.

- Lecturer Physical Presence vs. No Lecturer: The comparison between a lecturer's physical presence and a no lecturer's presence highlights the cognitive challenges posed by having a visible instructor. The results showed no significant difference in pupil dilation, suggesting that added visual stimuli do not universally increase cognitive load. However, the physical lecturer condition had a slightly higher mean pupil dilation, implying that processing non-verbal cues and instructional content may be more mentally demanding. This aligns with cognitive load theory, which states that extraneous cognitive load, such as processing non-essential visual elements, can interfere with learning (Sweller 2010).
- Lecturer Physical Presence vs. Avatar: The physical lecturer condition resulted in a significantly higher cognitive load than the avatar condition, underscoring the benefits of using avatars in video-based learning. Avatars do not increase cognitive load like physical instructors because they use simplified, purposeful gestures and lack dynamic facial expressions, reducing visual distractions. A physical lecturer's dynamic and complex visual cues increase cognitive effort, potentially distracting from the content. Conversely, avatars provide a simplified, controlled representation, reducing extraneous cognitive load and allowing more focus on learning (Alemdag 2022).
- No Lecturer vs. Avatar: A significant difference was found, with no lecturer presence leading to higher pupil dilation. The lack of instructional presence may impose extra cognitive demands due to missing guidance. In contrast, avatars provide a structured instructional presence while maintaining low visual complexity, helping mitigate cognitive load. Their simplified representation shifts cognitive load towards germane processing, enhancing learning outcomes (Sweller 2020). This supports findings that avatars lower unnecessary cognitive strain, which is vital in online education (Schöbel, Janson & Mishra 2019).
- Fixation Transition: The analysis of pupil dilation and attention shifts between instructional content (slides) and instructor presence suggests a nuanced effect on cognitive load (Kolnes, Uusberg & Nieuwenhuis 2024). With a physical lecturer, a small negative correlation suggests frequent attention shifts may help distribute cognitive effort, reducing strain. However, this pattern was absent in the avatar condition, where a small positive correlation suggests that transitions slightly increased cognitive load.

The findings highlight distinct effects of instructor presence on cognitive load. A physical lecturer induces the highest extraneous cognitive load, likely due to complex visual and non-verbal cues. In contrast, avatars reduce distractions, maintaining instructional presence while promoting germane cognitive load, which enhances learning. The absence of an instructor lowers cognitive load but limits learning gains, emphasising the need for an instructional presence in video-based learning, especially for complex materials.

Fixation transitions between slides and instructor presence varied by condition. With a physical lecturer, frequent transitions helped distribute cognitive effort, reducing cognitive strain. In contrast, in the avatar condition, transitions slightly increased cognitive load but also contributed to germane processing by aiding content assimilation. The findings confirm that instructor presence significantly influences cognitive load, with physical lecturers inducing the highest extraneous load, while avatars provide a more balanced learning experience (RQ1).

4.2 Relationship Between Cognitive Load and Learning Gain

The relationship between cognitive load and learning gain in video-based learning (VBL) is complex and context-dependent, as shown by pupil dilation analyses. The correlation between pupil dilation (a proxy for cognitive load) and learning gain varied across experimental conditions, suggesting that the instructor's presence influences how cognitive load affects learning outcomes.

In the no lecturer condition, a small positive correlation suggests that the absence of a presenter allows learners to allocate more resources to processing, enabling focused interaction with content and better learning outcomes. However, (Rodemer, Karch & Bernholt 2023) noted that the lack of instructional guidance might hinder comprehension of complex material due to missing social and contextual cues.

The physical presence of a lecturer showed a moderate negative correlation between pupil dilation and learning gain, indicating that a higher cognitive load may detract from learning. Processing visual and non-verbal cues can overwhelm learners, creating extraneous cognitive load, especially with challenging topics. This highlights the drawbacks of excessive cognitive load, as noted by (van der Wel & van Steenbergen 2018).

In the avatar condition, a small to moderate positive correlation suggests that avatars provide instructional presence while minimising extraneous demands. Simplified visual stimuli encourage optimal germane cognitive effort, positively influencing learning outcomes. Avatars can also capture attention and support learning by eliciting physiological responses (Ricou et al. 2024).

Comparing high and low-learning gain groups supports these findings. Low-performing learners in the physical lecturer condition exhibited higher pupil dilation, indicating that excessive cognitive load hindered information retention. No significant differences were found between high and low performers in the no-lecturer or avatar conditions, suggesting that both maintain cognitive load at manageable levels.

The findings show that cognitive load and learning gain are closely tied to instructor presence, with avatars striking a balance by reducing unnecessary strain while maintaining instructional cues. These results highlight the importance of instructional design in moderating cognitive load for effective digital education. Excessive cognitive load (e.g., physical lecturer presence) hinders learning, while balanced cognitive load (e.g., avatar presence) enhances outcomes (RQ2).

Avatars likely shift cognitive processing towards germane load, helping learners focus on content over distractions. This aligns with cognitive load theory, which emphasises optimising instructional design. Avatars should feature simple, non-distracting designs, avoiding dynamic elements that increase cognitive demands. Purposeful gestures, such as pointing, can direct attention to critical information and complement verbal content. Synchronising avatar behaviours with instructional material ensures coherence and reduces cognitive effort. Slower gestures and speech patterns can improve comprehension of complex topics. Subtle social presence, like consistent eye contact, fosters connection without overwhelming learners. These design considerations align with the study's findings, emphasising avatars' potential to balance cognitive load and instructional presence, providing an effective alternative to physical lecturers in VBL environments. In light of these findings, the study carries important implications for designing digital instructional environments, particularly as video-based learning becomes a standard in higher education and professional training. The evidence that avatar-based instructors can reduce extraneous cognitive load while maintaining learning effectiveness suggests a promising direction for scalable and cost-effective online education.

4.3 Limitations

While this study provides valuable insights into the relationship between instructor presence and cognitive load in video-based learning, several limitations should be acknowledged. While this study used pupil dilation as a proxy for cognitive load, it is important to acknowledge that other factors beyond cognitive effort can influence this measure. Emotional states, such as stress, can also cause pupil dilation (Heimerl et al. 2022), potentially confounding interpretations related solely to cognitive processing. Additionally, individual biological differences, including baseline pupil size or sensitivity to light, may affect how pupil responses manifest across participants. However, a within-subjects design helps control some of these variations.

Furthermore, using a single eye-tracking measure without triangulation from other physiological data may limit the comprehensiveness of cognitive load assessment. Future studies could integrate additional indicators such as heart rate variability or EEG to strengthen validity.

5. Conclusion

This study provided insights into the impact of instructor presence, including no lecturer, physical lecturer, and avatar, on cognitive load and learning outcomes in video-based learning. Eye-tracking in this study provided objective and real-time insights into the cognitive load, reducing the biases often inherent in self-reported measures. Eye-tracking data revealed that a physical lecturer increased extraneous cognitive load, negatively affecting lower-performing learners. In contrast, avatars balanced cognitive load by reducing distractions and enhancing germane cognitive processing. The absence of an instructor moderated cognitive load but limited learning gains, highlighting the importance of instructional presence for complex materials. These findings support avatars as an effective alternative, particularly in remote and asynchronous education.

Avatars offer a scalable and cost-effective solution for video-based learning, especially in MOOCs and remote settings. Their controlled gestures and visual design enhance germane processing and improve learning outcomes. To implement avatars effectively in diverse educational contexts, designers should prioritise simplicity, clarity, and alignment with instructional goals. Avatars should use minimal yet purposeful gestures and maintain a consistent visual presence without overwhelming motion.

By integrating the principles of Cognitive Load Theory, this research contributes to developing scalable, inclusive, and impactful e-learning technologies. Continued efforts to explore innovative design strategies and broaden the investigation's scope will help fully realise the potential of avatars and other digital tools in improving learning experiences across diverse educational contexts. Further research should explore dynamic avatars and integrate additional physiological measures like heart rate variability or electroencephalography to deepen the understanding of cognitive, emotional, and biological influences on learning.

Acknowledgements

We thank the Indonesian Education Scholarship, the Centre for Higher Education Funding and Assessment, and the Indonesian Endowment Fund for Education for the support of this research.

AI Statement: We declare that this paper is the original result of our research. We acknowledge the use of AI for improving grammar and paraphrasing. However, all edits were reviewed and verified by the authors in the final stage of the paper.

Ethics Statement: This study adhered to the ethical guidelines of Queen Mary University of London (approval number: QMERC20.565.DSEECS23.036). Participants provided informed consent, and data were pseudonymised.

References

Alemdag, E., 2022, 'Effects of instructor-present videos on learning, cognitive load, motivation, and social presence: A meta-analysis', *Education and Information Technologies*, 27(9).

Alhazmi, K., 2024, 'The Effect of Multimedia on Vocabulary Learning and Retention', *World Journal of English Language*, 14(6), 390.

Ayres, P. & Paas, F., 2007, 'Making instructional animations more effective: a cognitive load approach', *Applied Cognitive Psychology*, 21(6), 695–700.

Beege, M., Schroeder, N.L., Heidig, S., Rey, G.D. & Schneider, S., 2023, *The instructor presence effect and its moderators in instructional video: A series of meta-analyses*, *Educational Research Review*, 41.

Bettiol, S., Psereckis, R. & MacIntyre, K., 2022, 'A perspective of massive open online courses (MOOCs) and public health', *Frontiers in Public Health*, 10.

Bourguet, M.L., Xu, M., Zhang, S., Urakami, J. & Venture, G., 2020, *The Impact of a Social Robot Public Speaker on Audience Attention*, *HAI 2020 - Proceedings of the 8th International Conference on Human-Agent Interaction*.

Brünken, R., Seufert, T. & Paas, F., 2010, 'Measuring cognitive load', *Cognitive Load Theory*.

Chen, S. & Epps, J., 2014, 'Using task-induced pupil diameter and blink rate to infer cognitive load', *Human-Computer Interaction*, 29(4).

Chi, X., 2023, 'The Influence of Presence Types on Learning Engagement in a MOOC: The Role of Autonomous Motivation and Grit', *Psychology Research and Behavior Management*, Volume 16, 5169–5181.

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

Gilzenrat, M.S., Nieuwenhuis, S., Jepma, M. & Cohen, J.D., 2010, 'Pupil diameter tracks changes in control state predicted by the adaptive gain theory of locus coeruleus function', *Cognitive, Affective, & Behavioral Neuroscience*, 10(2), 252–269.

Heidig, S., Beege, M., Rey, G.D. & Schneider, S., 2024, 'Instructor presence in instructional videos in higher education: three field experiments in university courses', *Educational technology research and development*.

Heimerl, A., Becker, L., Schiller, D., Baur, T., Wildgrube, F., Rohleder, N. & Andre, E., 2022, *We've never been eye to eye: A Pupillometry Pipeline for the Detection of Stress and Negative Affect in Remote Working Scenarios*, *Proceedings of the 15th International Conference on PErvasive Technologies Related to Assistive Environments*, 486–493, ACM, New York, NY, USA.

Hessels, R.S., Benjamins, J.S., Cornelissen, T.H.W. & Hooge, I.T.C., 2018, 'A validation of automatically-generated areas-of-interest in videos of a face for eye-tracking research', *Frontiers in Psychology*, 9(AUG).

Kolnes, M., Uusberg, A. & Nieuwenhuis, S., 2024, 'Broadening of attention dilates the pupil', *Attention, Perception, & Psychophysics*, 86(1), 146–158.

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).

Lang, A., 2006, 'Using the limited capacity model of motivated mediated message processing to design effective cancer communication messages', *Journal of Communication*, 56(SUPPL.).

Leppink, J., Paas, F., Gog, T. van, Vleuten, C.P.M. van der & Merriënboer, J.J.G. van, 2014, 'Effects of pairs of problems and examples on task performance and different types of cognitive load', *Learning and Instruction*, 30.

Lim, J.Z., Mountstephens, J. & Teo, J., 2020, *Emotion recognition using eye-tracking: Taxonomy, review and current challenges*, *Sensors (Switzerland)*, 20(8).

Marx, J.D. & Cummings, K., 2007, 'Normalized change', *American Journal of Physics*, 75(1), 87–91.

Mayer, R.E., 2009, *Multimedia Learning*, Cambridge University Press.

Negi, S. & Mitra, R., 2020, 'Fixation duration and the learning process: an eye tracking study with subtitled videos', *Journal of Eye Movement Research*, 13(6).

Peysakhovich, V., Dehais, F. & Causse, M., 2015, 'Pupil Diameter as a Measure of Cognitive Load during Auditory-visual Interference in a Simple Piloting Task', *Procedia Manufacturing*, 3.

Pignatiello, G.A., Daly, B., Demaree, H., Moore, S. & Hickman, R.L., 2019, 'Comparing cognitive load levels among family members of the critically ill exposed to electronic decision aids', *Applied Nursing Research*, 50, 151192.

Polat, H., 2023, 'Instructors' presence in instructional videos: A systematic review', *Education and Information Technologies*, 28(7), 8537–8569.

Ricou, C., Rabadan, V., Mofid, Y., Aguillon-Hernandez, N. & Wardak, C., 2024, 'Pupil dilation reflects the social and motion content of faces', *Social Cognitive and Affective Neuroscience*, 19(1).

Rodemer, M., Karch, J. & Bernholt, S., 2023, 'Pupil dilation as cognitive load measure in instructional videos on complex chemical representations', *Frontiers in Education*, 8.

Rohrer, P.L., Delais-Roussarie, E. & Prieto, P., 2020, 'Beat Gestures for Comprehension and Recall: Differential Effects of Language Learners and Native Listeners', *Frontiers in Psychology*, 11.

Sablić, M., Mirosavljević, 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(4).

Sáiz-Manzanares, M.C., Marticorena-Sánchez, R., Martín Antón, L.J., González-Díez, I. & Carbonero Martín, M.Á., 2024, 'Using Eye Tracking Technology to Analyse Cognitive Load in Multichannel Activities in University Students', *International Journal of Human–Computer Interaction*, 40(12), 3263–3281.

Schöbel, S., Janson, A. & Mishra, A.N., 2019, *A configurational view on avatar design - The role of emotional attachment, satisfaction, and cognitive load in digital learning*, *40th International Conference on Information Systems, ICIS 2019*.

Serdar, C.C., Cihan, M., Yücel, D. & Serdar, M.A., 2021, 'Sample size, power and effect size revisited: simplified and practical approaches in pre-clinical, clinical and laboratory studies', *Biochemia medica*, 31(1), 27–53.

Souchet, A.D., Philippe, S., Lourdeaux, D. & Leroy, L., 2021, *Measuring Visual Fatigue and Cognitive Load via Eye Tracking while Learning with Virtual Reality Head-Mounted Displays: A Review*, *International Journal of Human-Computer Interaction*.

Sweller, J., 2010, *Element interactivity and intrinsic, extraneous, and germane cognitive load*, *Educational Psychology Review*, 22(2).

Sweller, J., 2020, 'Cognitive load theory and educational technology', *Educational Technology Research and Development*, 68(1).

Sweller, J. & Chandler, P., 1994, 'Why Some Material is Difficult to Learn', *Cognition and Instruction*, 12(3).

Sweller, J., Merrienboer, J.J.G. Van & Paas, F.G.W.C., 1998, 'Cognitive Architecture and Instructional Design', *Educational Psychology Review*, 10(3).

Tarchi, C., Zaccoletti, S. & Mason, L., 2021, 'Learning from text, video, or subtitles: A comparative analysis', *Computers & Education*, 160, 104034.

Wakefield, E., Novack, M.A., Congdon, E.L., Franconeri, S. & Goldin-Meadow, S., 2018, 'Gesture helps learners learn, but not merely by guiding their visual attention', *Developmental Science*, 21(6).

Wel, P. van der & Steenbergen, H. van, 2018, *Pupil dilation as an index of effort in cognitive control tasks: A review*, *Psychonomic Bulletin and Review*, 25(6).

Zagermann, J., Pfeil, U. & Reiterer, H., 2016, *Measuring cognitive load using eye tracking technology in visual computing*, *ACM International Conference Proceeding Series*, vols 24-October-2016.

Zhang, S., Wu, Y., Fu, Z., Lu, Y., Wang, Q. & Mingxuan, L., 2020, 'Psychometric properties of the Chinese version of the instrument for measuring different types of cognitive load (MDT-CL)', *Journal of Nursing Management*, 28(2).

Zheng, Z., Gao, S., Su, Y., Chen, Y. & Wang, X., 2022, 'Cognitive load-induced pupil dilation reflects potential flight ability', *Current Psychology*.