JOURNAL OF LEARNING ANALYTICS

Volume 11(3), 210–223. https://doi.org/10.18608/jla.2024.8235

# Is Seeing the Instructor's Face or Gaze in Online Videos Helpful for Learning?

Bertrand Schneider[1] and Gahyun Sung[2]

**Abstract**

Over the last decade, the prevalence of online learning has dramatically increased. As part of their curriculum, students are expected to spend more and more time watching videos. These videos tend to follow a widespread format: a screen recording of slides with a picture-in-picture (PiP) image of the instructor's face. While this format is ubiquitous, there is mixed evidence that it supports student learning. In this paper, we explore alternative formats for designing educational videos. Based on prior work showing the significance of joint attention for social learning, we create instructional videos augmented with the instructor's gaze and/or face. Testing these formats in a semester-long online course using a 2x2 experimental design, we found that showing the instructor's face had no significant effect on learning, while adding the instructor's eye-tracking data to the video promoted conceptual understanding of the material. Mediation analysis showed that joint visual attention played a significant mediatory role for learning. We conclude by discussing the implications of these findings and formulate recommendations for designing learning videos.

**Notes for Practice**

- Instructional videos are becoming increasingly prevalent. However, there is a lack of research on how different formats affect learning. This paper contrasted the effect of adding the instructor's gaze and/or face to videos during a semester-long course. Findings suggest that the instructor's face had no significant effect, while adding the instructor's eye-tracking data to the video promoted conceptual learning.
- By analyzing the eye-tracking data, we found that joint visual attention between the teacher and students was a significant mediator for learning. This means that teachers should carefully cue learner attention to important visual information when designing instructional videos.
- One implication for practice is that sensor data (such as eye tracking) has the potential to both capture learning processes and support them by making invisible information visible.

**Keywords:** Online learning, instructor presence, shared gaze visualizations, multimodal learning analytics

**Submitted:** 20/09/2023 — **Accepted:** 08/09/2024 — **Published:** 25/12/2024

*Corresponding author [1]Email: bertrand_schneider@gse.harvard.edu Address: Harvard University, Graduate School of Education 13 Appian Way, Longfellow Hall 333, Cambridge, MA, 02138, USA. ORCID iD: https://orcid.org/0000-0003-0922-2593*
*[2]Email: gcsung@uiowa.edu Address: University of Iowa, College of Education, 240 S Madison St., Iowa City, IA, 52242, USA. ORCID iD: https://orcid.org/0000-0003-0907-1377*

## 1. Introduction

In the last decade, educational videos have become one of the most prevalent forms of learning. The democratization of free video publishing has allowed anyone to upload instructional videos that are accessible worldwide. This movement was encouraged by charismatic educators and popular tutoring websites (e.g., Khan Academy), who popularized the idea that anyone can be an online teacher. Additionally, schools and companies started creating Massive Online Open Courses (MOOCs) where courses were made accessible at scale. At the same time, the popularity of "new" pedagogical approaches such as the flipped classroom model encouraged teachers to use pre-recorded videos for students to watch at home and focus class time on discussion and problem-solving. These developments mean that teaching through videos is no longer a niche area of education. Teachers, professors, professionals, and hobbyists across domains are now recording themselves to reach learners worldwide. The global COVID-19 pandemic has accelerated these changes and made learning from pre-recorded videos the norm instead of the exception.

A common video design found on platforms like YouTube© follows the "Picture-in-Picture" (PiP; or "stamp the instructor's face on a series of slides") format. One advantage of this format is that it is easy to generate with a standard laptop and webcam, and thus easily accessible to teachers worldwide; one disadvantage is that research suggests that this format does not enhance learning and might increase distraction and cognitive load (e.g., Wermeskerken et al., 2018; Kizilcec et al., 2014, 2015) This disconnect between research and practice has important implications for the millions of learners who spend hours every day learning from online platforms. We believe that there is much room for improvement when designing educational videos: it is a rich design space where more could be done to support learners' cognitive and affective engagement with the material taught. In short, we need more research to study the effect of different video augmentations and create new formats that go beyond recording the face of the instructor.

In this paper we propose to augment instructional videos with sensor data intended to enhance comprehension of the material taught. This approach leverages methods from a new field of research called Multimodal Learning Analytics (MMLA; Blikstein & Worsley, 2016), where high frequency sensors are used to support and assess learning. High frequency sensors refer to data collection tools that allow researchers to collect a large number of data points per second; for example, eye-trackers can capture gaze data between 30 and 120 times per second; or electrodermal data can be captured with devices such as the Empatica (Garbarino et al., 2014) and generate four data points per second. MMLA is becoming a popular methodology because of increasingly accessible sensing technology (such as affordable eye tracking, motion sensing, emotion detection, speech analysis, and physiological data collection tools) that allows researchers to capture fine-grained process data to complement scarce outcome measures, such as quizzes or tests (Schneider et al., 2024). More recently, computer vision algorithms have allowed researchers to collect data directly from video streams, for example about facial expressions (OpenFace; Baltrušaitis et al., 2016) or body pose (OpenPose; Cao et al., 2017). Of particular interest, eye-tracking technology allows us to observe how people communicate information through gaze (D'Angelo & Schneider, 2021), a key nonverbal mode of communication. Thus, our research questions are about understanding the effect of adding the instructor's face or gaze to instructional videos, and how this might facilitate or hinder learning. We also investigate the use of multimodal predictors for learning.

Subsequent sections are structured as follows. First, we review prior work on augmenting videos with the face or the gaze of an instructor. Second, we describe our study and interventions. We implemented these videos in a semester-long course taken by 52 university students during the COVID-19 pandemic. Third, we analyze the effect of the augmenting videos with facial or gaze information from the teacher on various learning dimensions. Fourth, we analyze student eye-tracking data and use joint visual attention as a predictor and assess its mediatory effect on learning. Finally, we summarize our findings and discuss implications for designing instructional videos that go beyond the PiP format.

## 2. Literature Review: The Challenges of Designing Learning Videos

Online learning has shown considerable growth over the last decade. Particularly, the COVID-19 pandemic has forced many institutions to switch to e-learning strategies. This shift was rarely viewed positively by students and teachers (Al-Mawee et al., 2021) and was found to be even more detrimental to learners in developing countries (Adnan & Anwar, 2020), students with disabilities (Denisova et al., 2020) and non-native English speakers (Hartshorn & McMurry, 2020). A major constituent of the online learning experience in both formal and informal learning environments is the asynchronous video lecture. These videos have effectively replaced many traditional in-person lectures, and students are now expected to spend increasing amounts of time watching them. Prior work suggests that the common PiP format does not have strong empirical evidence for its effectiveness (see section below on the "Instructor Presence Effect"). Thus, it is important to rethink how we deliver educational material online, especially through instructional videos. This paper investigates alternative video formats augmented with informational cues (other than the instructor's face). Below, we review the literature on Multimedia Principles for Designing Instructional Videos, the PiP format, then our suggested alternative format (i.e., videos with instructor gaze). Next, we discuss research on using multimodal data for educational research in order to contextualize the multimodal metrics used in the current study.

### 2.1. Multimedia Principles for Designing Instructional Videos

Mayer's (2005) multimedia theory offers several principles for designing effective instructional videos. These principles can explain why some video augmentations support learning more than others. Most importantly, any extraneous information that increases cognitive load and does not facilitate conceptual understanding should be avoided. Applied to instructional videos (Mayer, 2021), the Coherence Principle states that any irrelevant material should be removed to avoid overwhelming the learner. Additionally, the Redundancy Principle suggests that adding on-screen text that simply duplicates the narration can overload the learner's cognitive processing. Mayer also mentions that adding an instructor's image might not necessarily enhance learning (Image Principle); however, using a human voice rather than a machine voice can be more engaging for

ISSN 1929-7750 (online). 

learners. For a more recent systematic review of video design principles, see Fyfield et al. (2022).

Of interest to this study is the Signalling Principle: "Teachers should highlight important information to guide learners through the content." This can be achieved with visual cues or markers. As such, any information that can orient learners toward relevant content can help learning. Stull et al. (2018) provide some examples of signalling interventions: teachers can draw graphics in real time, instead of referring to already drawn graphics (Dynamic Drawing Principle); they can shift gaze between the audience and a board to cue learner attention (Gaze Guidance Principle); or they can shoot videos from a first-person perspective instead of a third-person perspective (Perspective Principle). Several studies have found that signalling can support learning: for example, in a narrated animation explaining how airplanes take off, teachers can emphasize some terms (a type of verbal signalling); Mautone and Mayer (2001) found that this improved transfer compared to students in a control group. Another type of signalling is using pointing gestures to guide the learner's visual attention; Li et al. (2019) found that this improved retention and transfer immediately after the lesson and after a one-week delay. These principles inspired the interventions of this study, where the eye-tracking data of the teacher is used to augment instructional videos.

### 2.2. Instructor Presence Effect: Augmenting Learning Videos With the Instructor's Face (PiP)

One widespread assumption is that seeing an instructor is beneficial to learning. There are several reasons why: seeing the instructor is supposed to increase social presence and motivation, which is hypothesized to have a positive effect on learning (Alemdag, 2022). A more practical reason is technical access: every computer is equipped with a webcam, and there are many software programs that can overlay a person's face onto a screen recording. Finally, teachers have a penchant to replicate in-person instruction, for example where lecturing plays a predominant role.

However, the advantages of doing so are mixed. Several recent literature reviews on instructor presence conclude that showing the face of an instructor has a null or mixed effect on learning. Alemdag (2022) analyzed 20 studies and found no significant effect of instructor presence on learning but found that it increased cognitive load and motivation. Henderson and Schroeder (2021) looked at 12 studies and found no evidence that an instructor should be included on instructional videos, besides the fact that some studies reported increased student satisfaction when the instructor was present. While some argue for the inclusion of the teacher's face (Paciej-Woodruff, 2021) supported by studies finding that learning performance is improved with PiP or lecture recordings compared to Khan-style voice-over videos (Chen & Wu, 2015; Kokoç et al., 2020), others find that it has no significant effect on retention (Ng & Przybyłek, 2021) or that it attracts learner attention at the expense of what is explained (van Wermeskerken et al., 2018). Indeed, studies find that an instructor's face attracts up to 40% of learner attention if present in a video (Kizilcec et al., 2014), or that a human face deters the learning of information nearest to the face (Djamasbi et al., 2012). In other cases, studies have found that while students like seeing the instructor's face, it does not enhance comprehension (Kizilcec et al., 2014, 2015; Wilson et al., 2018). In sum, the literature casts doubt on the fact that instructor presence is beneficial to learning, prompting the investigation of alternative formats.

### 2.3. Shared Gaze Visualizations: Augmenting Learning Videos With the Instructor's Gaze

Developmental and educational theories show that joint attention plays a significant role in learning and teaching. Joint attention (JA) refers to the shared focus of two individuals on an object or a topic. It is achieved when one individual alerts another to an object by means of eye-gazing, pointing, or other verbal or nonverbal indications. Joint attention is considered a critical component in social development, language acquisition, and cognitive understanding (Tomasello, 1995). Humans need joint attention to coordinate their actions with others and to learn from them. From children acquiring their first words, teenagers learning from schoolteachers, students collaborating on a project, to any group of adults working toward a common goal, joint attention is a fundamental mechanism for establishing a common ground between individuals (Clark & Brennan, 1991). A good common ground ensures that group members refer to the same objects, locations, facts, concepts, and ideas. It is a fundamental building block for effective communication between human beings, even more so in situations where skills or knowledge are exchanged. Autistic children, for instance, are known for lacking the ability to coordinate their visual attention with their caregivers, which is associated with many social impairments, including uneven language development and learning disabilities (Mundy et al., 1990). The importance of joint visual attention (JVA) in learning has been demonstrated for young adult learners both qualitatively (Barron, 2003) and quantitatively (e.g., using eye tracking; Schneider & Bryant, 2024). The results have inspired various interventions to support collaboration and learning, for example through shared gaze visualizations (SGVs). SGVs are either live or pre-recorded eye-tracking visualizations that indicate what an individual is looking at, aiming to facilitate referencing, disambiguate utterances, and improve mutual understanding (for a review of the benefits of SGVs, see Angelo & Schneider, 2021).

For pre-recorded videos, research shows that Eye Movement Modelling Examples (EMME) have the potential to support integrative processing of different representations (such as text and images) to facilitate recall and transfer (Mason et al., 2015). A recent review article finds that EMMEs lead to a net performance gain ($d = 0.43$; Xie et al., 2020). Specifically, gaze augmented videos have been found to aid learners in attending faster and longer to task-relevant materials, minimizing the

likelihood of miscommunication, and letting them follow the thought process of an instructor (Pi et al., 2020; Melnyk et al., 2021). These effects can vary across task and learner contexts; for instance, while it can be helpful for conducting visual searches in complex diagrams (Jarodzka et al., 2013), it can be distracting when gaze wanders without a clear intention to signal (D'Angelo & Gergle, 2016). Similarly, SGVs have been shown particularly to help low-achieving learners who need additional cues from instructors (D'Angelo & Schneider, 2021).

EMME has been applied to various settings, such as MOOCs. For example, Sharma et al. (2015) found that showing the gaze of the teacher made the video content easier to follow. Sharma et al. (2016) designed a feedback tool to increase joint visual attention with the teacher when it fell below a certain threshold and found that it significantly increased learning gains. Sharma, D'Angelo, et al. (2016) augmented video recordings on cloud identification (e.g., stratus, cumulus, cirrus) with the teacher's gaze or a pointer, and found that SGV increased learning gains compared to no visual aid (but not with the pointer). In more traditional settings, Špakov et al. (2019) designed a two-way gaze sharing system in a tutor-tutee situation; they found that students liked the tutor's gaze marker during exercises but found it distracting during slide reading. The tutor used the student gaze point to see what they were focusing on when helping them individually. Sauter et al. (2022) replaced eye-tracking data with a pointer manipulated by the teacher; they found that the distance between student gaze and the pointer predicted student learning.

While SGVs show promise, they do not guarantee that learners process the explanations of instructors just because they are cued to follow their gaze. For example, Stull et al. (2018) found that gaze guidance cues did not increase student learning or engagement. As such, continued research is needed to optimize the design and deployment of SGVs in instructional videos for widespread use.

### 2.4. New Methods for Assessing Online Learning from Videos

While multimodal data streams can augment video recordings (for example by overlaying the eye-tracking data from the instructor onto a video), they can also provide new ways to compute learning indicators, for example by analyzing the eye-tracking data from students. This is important because the most traditional way of assessing learning from videos is to administer short quizzes or learning tests. This kind of measure is a limited one-time assessment of student learning. Recently, educational researchers have started to pay more attention to *process* data (Schneider, 2023), especially how it can be automatically collected from high frequency sensors, for example by capturing what learners pay attention to (eye-tracking data; Schneider, 2020), their affective state (electrodermal data; Schneider et al., 2020), or their bodily movement (motion data; Blikstein & Worsley, 2016); for a review on how sensors can capture social interactions, see Schneider et al. (2022). For this reason, we are interested in exploring alignment measures between students and the instructor. Based on our prior work, there are reasons to believe that joint visual attention might be correlated with student learning (Sharma, Jermann & Dillenbourg, 2014; Schneider & Bryant, 2020). While JVA has often been explored in collaborative contexts (e.g., Chen et al., 2021; Guo & Barmaki, 2020) and teaching (e.g., Sung et al., 2021), we are interested in computing the extent to which students align their attention with the visualized gaze of the instructor in the current setting. We believe that this measure would not only capture whether students are following along but also processing information like the instructor. In our study, we collected eye-tracking data from student webcams as they were watching the videos and computed JVA with the instructor. See section 4.6 below for more information.

### 2.5. General Description of the Study and Contributions

To our knowledge, no study has attempted to directly compare SGV and the Instructor Presence Effect in a semester-long course. Most prior work was lab-based and collected data on a shorter time frame. Additionally, other studies have primarily relied on outcome measures (e.g., learning tests) and few of them have collected fine-grained process-data. Finally, we did not find any comparable study that has collected and assessed the predictive value of webcam-based attentional measures for learning.

Given these gaps in the literature, we assessed the benefits of SGVs and the instructor's face on instructional videos. Students (N=52) enrolled in a semester-long course watched weekly videos on using quantitative data in educational research and took a weekly quiz to measure their learning. While students were watching the instructional videos, we also collected web-based eye-tracking data on them. We assess the effects of these interventions on learning and investigate whether they had an impact on joint visual attention with the instructor.

The contributions of this paper are as follows. First, we augment online learning videos with multimodal information (i.e., with the instructor's gaze and/or face) and assess the effect of this intervention on learning. Second, we explore the usefulness of webcam-based sensing technologies for deriving indicators of student learning. Most prior work used dedicated hardware for collecting eye-tracking data; in this project, we investigate whether webcam-based data can yield useful metrics for predicting learning. This is a key factor to consider if we are to scale up subsequent research. Third, we compute measures of joint visual attention between students and the instructor and assess the mediatory effect of this measure on learning.

Finally, we provide some preliminary design principles for augmenting videos with gaze data so that our approach might be replicated in research or practice. We conclude by discussing the implications of our findings and conclude with recommendations for designing online videos.

## 3. Research Questions

Based on this literature review, our main research questions (RQs) are as follows:

RQ1. Do videos augmented with the instructor's face (Instructor Presence) and/or gaze (Shared Gaze Visualizations) have a positive effect on student learning?

RQ2. Does joint visual attention mediate learning?

## 4. Methods

### 4.1 Participants

Participants (N=52) were graduate students at a private graduate school of education in the Northeastern region of the U.S. enrolled in an introductory quantitative data analysis course. Fifty-four percent were female and 46% male. Most students were in their twenties. Before enrolling in the course, students were asked to complete a brief interest and skills assessment survey. Students indicated having little to no experience with data mining techniques (69%), or psychometrics (79%); in contrast, they indicated having some or strong experience (69%) with learning theories such as constructivism or constructionism. Student interest in the class limited participant recruitment. We note that our sample size is similar to that of researchers who used a similar multi-conditions experimental design (e.g., McAlpin et al., 2023). Later, we also conducted a post-hoc power analysis, using a type 1 error rate of 0.05 and observed effect sizes. This resulted in an estimated power of 77%, which is interpreted as a 77% likelihood to detect an existing effect. This does not exceed, but very closely approaches the conventional power threshold of 80% used in medical trials (Serdar et al., 2021).

### 4.2. Context

Participants were enrolled in a 13-week course in the Fall 2020 semester. Due to the pandemic, the course was taught exclusively online. The course structure involved real-time instruction, hands-on projects, and twelve weekly asynchronous videos. The videos were uploaded to an online learning platform that captured and locally processed participant webcam feeds as they watched videos to generate eye-tracking data. Participants were informed of the purposes of the recordings and had continuous access to their own data on the learning platform. All participants signed a consent form to agree to the data collection process.

### 4.3. Material

When recording the video lectures, a webcam captured the instructor's face and a Tobii 4C eye tracker captured his eye movements. Four versions of the videos were created: some of the videos had the instructor's face overlaid on top, and some were augmented with the instructor's gaze (see Table 1). In each video, the instructor tried to exhibit various affective states (e.g., enthusiasm, frowning when concepts were more difficult, etc.) to make the content as engaging as possible.

After each video, participants took a quiz testing their knowledge of the content of the video. There were ten items each week, for a total of 120 questions over the entire semester. The quizzes included questions about facts, concepts, procedures, and examples of applications of large datasets in education. For example, in weeks 6–7, where the topic was machine learning, conceptual questions included these: "What are potential solutions to a model underfitting the data?"; "What are potential solutions to a model overfitting the data?" Factual questions included "Which of the following algorithms are probabilistic?" Procedural questions included "What is the first step in the K-means algorithm?" Finally, application questions included "Which of the following are examples of applications of machine learning?" In total, there were 50 factual questions, 35 conceptual questions, 24 procedural questions, and 11 questions about applications. We double coded 20% of the questions and reached an agreement of 83% (Krippendorff's Alpha = 0.72), which indicates a "good" agreement between raters.

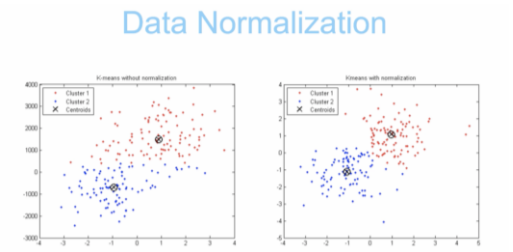
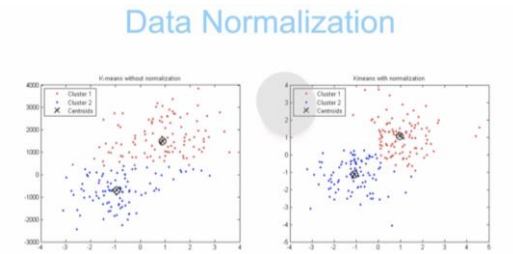
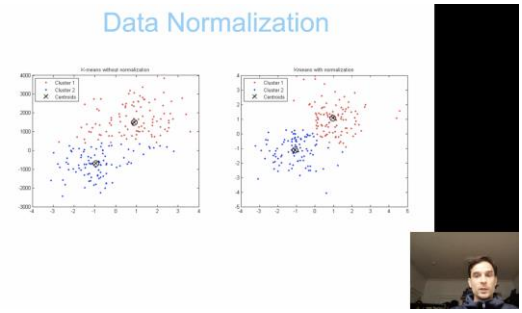
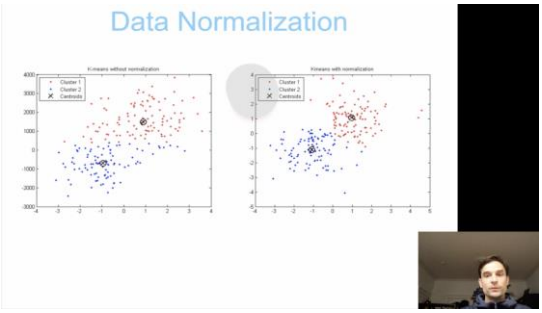
### 4.4. Design

We used a 2x2 experimental design to test the effect of the instructor's face and gaze on the videos: a quarter of the students saw the raw video; a quarter saw the instructor's face next to the video; a quarter saw the instructor's gaze on the video; and a quarter saw both (Table 1). We used a 2x2 design to test for an interaction effect between the two interventions: prior work has shown that seeing the gaze of an instructor is beneficial to learning (e.g., Mason et al., 2015), even though it increases cognitive load; however, we expected that adding the face would create too much cognitive load and be detrimental to learning, by having students monitor both the gaze and the face of the instructor.

To assess the effect of each intervention, we compared students who saw the videos with and without the instructor's gaze, and students who saw the same videos with and without the instructor's face. Participants were randomly assigned to one of the experimental conditions for the entire semester. While we cannot guarantee that students did not exchange information about their experimental conditions, this potential bias is mitigated by the fact that this was an online course and students never met in person.

While the sample size per cell could be larger (N=13), what matters most is the total sample size (N=52) because we are not interested in conducting comparisons between individual cells (Westfall, 2015). What we care about is the difference between control and treatment groups (e.g., with/without the face of the instructor), which means that we have 26 data points for each condition.

**Table 1:** Sample Snapshots of Instructor Gaze

| **Conditions** | No instructor's gaze (N=26) | With instructor's gaze (N=26) |
|---|---|---|
| No instructor's face (N=26) |   |   |
| With instructor's face (N=26) |   |   |

*Note: The table shows sample snapshots of a video in each of the 2x2 experimental conditions used in this study. Each row/column contained 26 participants. The grey circle on the right column indicates the location of the instructor's gaze as seen by the students. In this example, student attention is drawn to the y-axis (which is crucial in understanding why data normalization leads to better clusters).*

### 4.5. Procedure

Each week, students watched an instructional video on the topics mentioned above. Each video was between 20 and 30 minutes long, and provided factual, procedural, and conceptual information. There was no time limit for watching the video, and students could watch it as many times as they wanted. When they felt ready, students took a 10-question quiz with multiple-choice answers. The time limit on the quiz was 15 minutes, to ensure that students used their own notes and understanding of the material to answer the questions. We found that the time limit minimized overreliance on the video. Students had until a specified deadline to complete the quiz, after which the correct answers were released.

### 4.6. Multimodal and Outcome Measures

Table 2 provides an overview of the measures used in this study. Independent variables (interventions and demographics; left column) are described in sections 4.1 and 4.4 Mediator (middle column) are described below. Dependent variables (right column) are described in section 4.3.

**Table 2:** Summary of the Variables Used in Analysis

| Independent variables | Mediator | Dependent variables |
|---|---|---|
| Intervention | Eye-tracking data | Learning outcomes (Quiz scores) |
| Presence/absence of gaze | Joint visual attention | 50 factual questions |
| Presence/absence of face | | 35 conceptual questions |
| | | 24 procedural questions |
| | | 11 questions about applications |

We captured the instructor's eye gaze using a Tobii 4C, and student gaze using WebGazer (Papoutsaki et al., 2016). WebGazer.js is an open-source, web-based eye tracking library that leverages common web technologies such as JavaScript, HTML, and CSS to provide real-time gaze prediction on the screen of a computer or mobile device. It uses data from a webcam to estimate where a user is looking on the screen. WebGazer improves its accuracy over time, by using user clicks at calibration data for estimating the x and y coordinate of user gaze on a screen. To compute joint visual attention, we compared the location of the instructor and student gaze for each video frame. We standardized gaze coordinates between 0 and 1 because the video dimensions varied based on student/instructor browser window size and resolution. We then computed the distances between the instructor's coordinates and each student's coordinates and assessed different thresholds to determine if they shared the same attentional focus. Because the resolution of webcam-based eye tracking is less accurate than with dedicated hardware, we considered larger thresholds than traditional eye-tracking studies (e.g., Schneider & Bryant, 2024). While we found similar trends across different threshold values, we took a more conservative approach and considered distances below 0.25 (i.e., a quarter of video size) between two gazes to count as joint visual attention. Table 3 provides a high-level overview of the data collected for each student and averaged over the entire semester (scores and synchrony measures are expressed in percentages), and Table 2 provides a summary of the variables considered in the analyses below.

**Table 3:** Head of the High-Level Data Frame

| | A | B | C | D | E | F | G | H | I |
|---|---|---|---|---|---|---|---|---|---|
| 1 | ParticipantID | gaze | face | fact_mean | concept_mean | procedure_mean | application_mean | total_mean | jva |
| 2 | 0 | 0 | 1 | 0.9125 | 0.852777778 | 0.796666667 | 0.777777778 | 0.834930556 | 0.027101559 |
| 3 | 1 | 1 | 0 | 0.654545455 | 0.651515152 | 0.566666667 | 0.777777778 | 0.662626263 | 0.020963584 |
| 4 | 2 | 1 | 1 | 0.938888889 | 0.901388889 | 0.84 | 0.777777778 | 0.864513889 | 0.048018655 |
| 5 | 3 | 0 | 1 | 0.708928571 | 0.772222222 | 0.746666667 | 0.833333333 | 0.765287698 | 0.234768426 |
| 6 | 4 | 1 | 1 | 0.701984127 | 0.630555556 | 0.576666667 | 0.555555556 | 0.616190476 | 0.17888953 |
| 7 | 5 | 1 | 1 | 0.854166667 | 0.830555556 | 0.866666667 | 0.777777778 | 0.832291667 | 0.095474523 |
| 8 | 6 | 0 | 0 | 0.9625 | 0.872222222 | 0.78 | 0.666666667 | 0.820347222 | 0.07545618 |
| 9 | 7 | 0 | 1 | 0.983333333 | 0.979166667 | 1 | 0.888888889 | 0.962847222 | 0.005025422 |
| 10 | 8 | 0 | 0 | 0.872222222 | 0.841666667 | 0.833333333 | 0.722222222 | 0.817361111 | 0.060829453 |
| 11 | 9 | 0 | 1 | 0.751190476 | 0.734722222 | 0.746666667 | 0.722222222 | 0.738700397 | 0.007235924 |
| 12 | 10 | 0 | 1 | 0.936111111 | 0.806944444 | 0.9 | 0.777777778 | 0.855208333 | 0.007394786 |
| 13 | 11 | 0 | 0 | 0.776388889 | 0.793055556 | 0.78 | 0.5 | 0.712361111 | 0.084305684 |
| 14 | 12 | 1 | 0 | 0.979166667 | 0.966666667 | 0.95 | 0.777777778 | 0.918402778 | 0.0413967 |
| 15 | 13 | 0 | 1 | 0.955555556 | 0.872222222 | 0.916666667 | 0.777777778 | 0.880555556 | 0.148417665 |
| 16 | 14 | 1 | 0 | 0.901388889 | 0.876388889 | 0.796666667 | 0.722222222 | 0.824166667 | 0.159593041 |

*Note: Head of the high-level data frame used to compare experimental conditions (presence of face, gaze) on quiz scores (fact, concept, procedure, application). The last column is a synchrony measure of the gaze (jva = joint visual attention).*

### 4.7. Data Analysis

In this section, we describe the measures we computed from the data and how we addressed our research questions. For RQ1, we used a between-subjects Multivariate Analysis of Variance (MANOVA) to assess the effect of the instructor's gaze and face overlaid on the weekly videos. Learning was measured through student scores on the quiz questions. For RQ2, we correlated joint visual attention with learning scores and assessed several mediation models.

## 5. Results

### 5.1. Exploratory Data Analysis

A preliminary step in our data analysis was to check the quality of the data collected while students were watching the videos. Since there were hardware, environmental, and individual differences between students, data quality tended to vary. We

computed students' joint visual attention with the instructor and plotted the data for the entire semester. We show six exemplar graphs in Figure 1 for the eye-tracking data. Each graph represents a student, and the colours represent the different weeks of the semester. The leftmost two graphs show "healthy" data with a sizable dataset for each week. The next two graphs each show "borderline" data sizes of around five thousand, suggesting that the data may not be comparable to that of other students. The next two graphs show unusable datasets due to missing data. The rightmost graph shows the number of data points for each student (i.e., each student is represented by a bar). We looked at a bar graph to identify "dips" in the data (i.e., when the slope significantly increased). In the case of the eye-tracking data, we used two thresholds for removing students: when they had less than 5k and 10k data points. We found a stronger correlation with a more conservative threshold (i.e., 10k), which we used for our analyses.

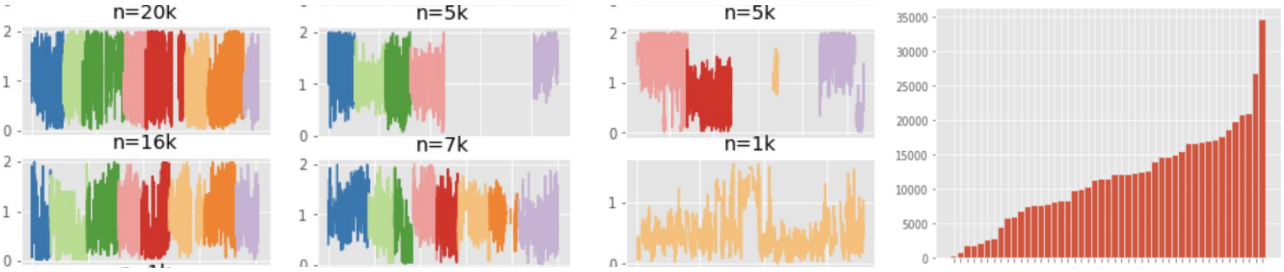



**Figure 1:** Left: eye-tracking data for six students (the distance with the instructor's gaze is shown on the y-axis; time is shown on the x-axis; colours represent the different weeks of the semester). Right: total number of data points for the semester (each bar represents a distinct student enrolled in the course).

### 5.2. RQ1: Do Videos Augmented With the Instructor's Face and/or Gaze Positively Affect Student Learning?

To answer the first research question, we aggregated student quiz scores over the entire semester (see Figure 2) in four categories (facts, concepts, procedures, applications) and assessed for significant differences using a MANOVA. Each variable was checked for normality, outliers, and homogeneity of variance. For testing homogeneity of variance, we used Leven's test (based on medians) and found that none of the results were significant. This means that across groups, quiz scores are considered to have comparable variance. For outliers, we generated boxplots (see Figure 2, left side) for each question category. Factual and procedural questions had a few outliers; we replicated our analyses below without them and found that they did not affect the significance of our results. We assessed normality using the Kolmogorov-Smirnov test and found that responses on procedural ($p < 0.001$) and application ($p < 0.05$) questions were significant (i.e., not normal). For these categories, we also conducted non-parametric tests.

The MANOVA revealed a significant effect of overlaying the gaze of the instructor on conceptual questions: $F(1,44) = 4.42$, $p = 0.04$, Cohen's $d = .77$ (no-gaze: *mean*=0.48, *SD*=0.27; visible-gaze: *mean*=0.67, *SD*=0.22), but not on other question types (factual, procedural, applications). We did not find any significant effect of seeing the instructor's face on any of the four question types, or any significant interaction effect between the two conditions ($F < 1$). Because some of our dependent variables were not normally distributed, we conducted the same analyses using a non-parametric test (Kruskal-Wallis H Test). The results were unchanged: only the gaze intervention had a significant effect on conceptual scores: $U(1) = 4.27$, $p < 0.05$.

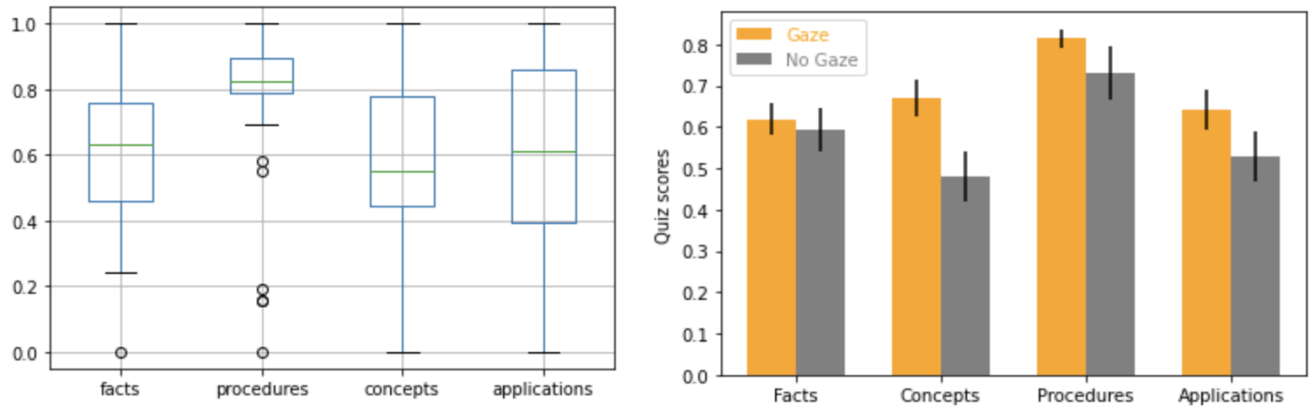



**Figure 2:** Left: boxplots of learning scores on the four subdimensions of the quizzes over the entire semester. Right: bar charts for the different experimental conditions (whiskers indicate standard errors).

ISSN 1929-7750 (online). 

To take the temporal dimension of our data into account (i.e., students were tested every week for an entire semester), we replicated the results above using a multilevel model in SPSS. Weeks and experimental conditions were modelled as fixed effects, and we looked at their effect on conceptual scores. We again found a significant effect of seeing the gaze of the instructor on conceptual questions: $F(1,531) = 4.12$, $p < 0.05$. There was no significant interaction effect between time (i.e., weeks) and experimental conditions: $F(1,531) = 2.5$, $p = 0.11$.

### 5.3. RQ2: Does Joint Visual Attention Mediate Learning?

First, we assessed whether joint visual attention was correlated with quiz scores. We found that joint visual attention was significantly correlated with conceptual scores: $r(24) = 0.48$, $p < 0.005$, but not other categories (fact, procedures, applications). Second, we assessed a mediation model (see Figure 3) to further delineate the significance of joint visual attention. The model examined the effect of the experimental conditions on conceptual learning, mediated by joint visual attention. Because of our small sample size, we used a bootstrapping with 1,000 replicates for uncertainty estimation. We mark a finding as significant if zero is outside the resulting 95% bootstrap confidence interval (Efron & Tibshirani, 1993).

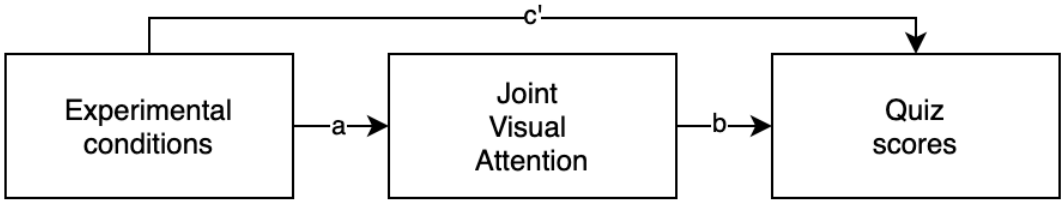



**Figure 3:** The mediation model assessed in this paper.

The effect of experimental conditions (i.e., the instructor's face or gaze) on joint visual attention was significant ($p<.01$, CI: [.012; .08]) as was the effect of JVA on conceptual questions ($p<.01$, CI: [.53; 3.34]). The total effect of the model was significant ($p<.05$, CI: [.047; 0.33]).

In summary, the correlation analysis shows that joint visual attention seems to be associated with higher quiz scores on conceptual questions. Additionally, the mediation analysis suggests, together, that joint visual attention is a significant mediator for the experimental conditions on learning. We further discuss these findings below.

## 6. Discussion

We compared the effect of seeing the instructor's gaze and face in an instructional video during a semester-long course. Students watched weekly videos and completed a quiz testing their understanding of the material. We collected eye-tracking data while they learned from the videos and explored the predictive value of joint visual attention.

Our first research question was about the effect of adding the instructor's face or gaze to the videos and how it impacted student learning. We found that adding the face of the instructor had no effect on learning, which is in line with prior work (Djamasbi et al., 2012). This is not to suggest that showing the instructor's face is without benefit; this may influence aspects of student learning experiences other than conceptual understanding, such as their enjoyment, long-term motivation, or rapport with the instructor (Kizilcec et al., 2014, 2015; Wilson et al., 2018). However, what effect it may have had did not translate into learning. Visualizing eye-tracking data, on the other hand, increased scores on conceptual questions (but did not affect other types of learning, such as learning about facts, procedures, or applications). This is also in line with prior work that has found that gaze augmented videos aid learners attend to task-relevant materials, minimize the likelihood of miscommunication, and help students follow the thought process of an instructor (Pi et al., 2020; Melnyk et al., 2021). There are several potential interpretations of this result, inspired by the *signalling effect* described by Mautone and Mayer (2001): The gaze might have served as a nonverbal cue that guides student attention toward critical elements or ideas in the lesson, enhancing conceptual understanding. Since facts and procedures can often be outlined or highlighted in text or verbally emphasized, the effect of gaze guidance may not be as pronounced for these types of learning, since student cognitive load might be lower in these situations. Additionally, the inclusion of the teacher's gaze might help manage cognitive load (Mayer, 2005; Mayer, 2021): conceptual understanding requires connecting new knowledge with existing cognitive structures, and effective management of cognitive load would particularly benefit this process. The memorization of facts or the learning of procedures might rely less on cognitive load management and more on repetition. Finally, gaze awareness can also "contribute to an improved feeling of social presence" (Akkil et al., 2018), making students feel more connected and thus more motivated to engage with conceptual material. The implication of these findings is that it might be more beneficial to add gaze data to learning videos, contrary to the customary practice of adding the teacher's face, if we primarily care about supporting student conceptual learning.

In our second research question, we explored the predictive value of eye-tracking data. Prior work suggests that shared attention tends to be linked to interactions of higher quality (Schneider et al., 2018), and increased learning gains (D'Angelo & Schneider, 2021). Indeed, we found in our dataset that joint visual attention was positively and significantly associated with student scores on conceptual questions. This suggests that students who paid attention to what the instructor was looking at tended to learn concepts better, which is in line with prior findings (Schneider et al., 2023). Additionally, a mediation analysis suggests that joint visual attention significantly explains the variance in learning, and similarly, JVA is a significant mediator for the association between experimental conditions and learning (as found in other studies by Schneider & Pea, 2013). This suggests that students were not absent-mindedly following the moving dot representing the gaze of the instructor on the video: establishing joint visual attention with the instructor was associated with higher learning scores.

While our results are mostly consistent with prior work, they generalize findings from shorter, controlled (lab) studies (e.g., Sharma et al., 2014; Jarodzka et al., 2013) to a course taught during an entire semester. We also assessed whether webcam-based eye-tracking measures could be used to compute measures of joint visual attention, which previously had been done with dedicated hardware (e.g., Schneider et al., 2023; Mason et al., 2015). Finally, this paper extends prior results by directly comparing the "instructor presence" with shared gaze visualizations and showing that the latter seems more beneficial to student learning.

### 6.1. Design Considerations

While we found a positive effect of sharing gaze data, we do not believe that it is a silver bullet for making online videos easier to understand. According to a review on SGVs (D'Angelo & Schneider, 2021) there is a tension between gaze supporting communication (e.g., by facilitating referencing) and distracting the viewer (e.g., by displaying fine-grained fixations or saccades). In the design of the videos presented in this study, the instructor had to consciously minimize scanning behaviours and maximize the use of his gaze as a communication medium. This was not intuitive for the instructor since we usually do not consciously control our eye movements. While there is time and energy gained by not having to manually annotate every slide with arrows and highlights for comparable effect, recording eye-tracking data adds cognitive load when recording a lesson. Consequently, instructors who know their material well and have been teaching it for several years may be the best candidates for creating high-quality learning videos augmented with gaze data. Additionally, particular concepts and formats might benefit more from SGVs than others (e.g., material involving complex representation). These design considerations are in line with best practices for designing educational videos (Mayer et al., 2020).

### 6.2. Limitations

There are several limitations to this study. First, our data cleaning steps include removing participants who did not have enough data. This reduced our sample size and introduced the risk that data may have not been missing at random (e.g., students who were less engaged spent less time on lessons). For these reasons, some of our findings should be replicated before any generalization claims can be made. Second, we acknowledge that webcam-based data collection tools (eye trackers) are not as accurate as solutions using dedicated hardware. This limits the accuracy of the data, especially across a variety of participants and settings (e.g., lighting, hardware). Lastly, related to this point, there are several unknown factors that can compromise the quality of the data. Different students had different hardware, sometimes more affordable laptops that are not suited for processing real-time webcam data. Sanity checks, such as the one discussed in the "Exploratory Data Analysis" section, will be continuously needed in future studies to discern the reliability of the data.

### 6.3. Future Work

For future work, we plan to replicate these results with a new cohort of students. This will increase confidence in the generalizability of our findings. Because recording the videos increases the instructor's cognitive load, we are also interested in developing better tools for recording their gaze (e.g., by changing the type of visualization used, or by letting the instructor turn it on and off when necessary). Alternatively, new implications may be found by providing the same features to students, so that they can customize the gaze visualization to their personal preferences (D'Angelo et al., 2019). Lastly, since prior work has identified effects of video interventions on learners' cognitive loads (Alemdag, 2022), this should be considered in future work.

## 7. Conclusion

Our results suggest that SGVs may be a promising way to enhance the traditional PiP video format in online learning for conceptual learning and for learners who need additional support in cognitive resources. Given that this is a result from one class with an idiosyncratic population, topic, and social setting, we do not argue that SGVs are panaceas; rather, we contend that there is a need to conduct future tests to validate their effects, focusing on certain areas found to be promising in the current study. Namely, our results suggest that SGVs are helpful in online learning when presenting complex, multiple

representations (for additional considerations, see D'Angelo & Schneider, 2021).

As an additional contribution, we find that webcam-based multimodal measures could be meaningful proxies of the learner's cognitive process. We found that joint visual attention is a particularly theoretically sound construct, as well as an empirically promising indicator of conceptual understanding. In sum, these findings further our understanding of the factors that contribute to effective online learning experiences and pave the way for studying innovative ways of augmenting videos with multimodal information.

## Declaration of Conflicting Interest

The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

## Funding

The authors declared no financial support for the research, authorship, and/or publication of this article.

## References

Adnan, M., & Anwar, K. (2020). Online learning amid the COVID-19 pandemic: Students' perspectives. *Journal of Pedagogical Sociology and Psychology*, *2*(1), 45–51. https://doi.org/10.33902/jpsp.2020261309

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

Al-Mawee, W., Kwayu, K. M., & Gharaibeh, T. (2021). Student's perspective on distance learning during COVID-19 pandemic: A case study of Western Michigan University, United States. *International Journal of Educational Research Open*, *2*, 100080. https://doi.org/10.1016/j.ijedro.2021.100080

Akkil, D., Thankachan, B., & Isokoski, P. (2018). I see what you see: Gaze awareness in mobile video collaboration. *Proceedings of the 2018 ACM Symposium on Eye Tracking Research & Applications* (ETRA '18), 14–17 June 2018, Warsaw, Poland (Article 32). ACM Press. https://doi.org/10.1145/3204493.3204542

Barron, B. (2003). When smart groups fail. *Journal of the Learning Sciences*, *12*(3), 307–359. https://doi.org/10.1207/s15327809jls1203_1

Blikstein, P., & Worsley, M. (2016). Multimodal learning analytics and education data mining: Using computational technologies to measure complex learning tasks. *Journal of Learning Analytics*, *3*(2), 220–238. https://doi.org/10.18608/jla.2016.32.11

Baltrušaitis, T., Robinson, P., & Morency, L.-P. (2016). OpenFace: An open source facial behavior analysis toolkit. *2016 IEEE Winter Conference on Applications of Computer Vision* (WACV), 7–10 March 2016, Lake Placid, NY, USA. IEEE. https://doi.org/10.1109/wacv.2016.7477553

Cao, Z., Simon, T., Wei, S.-E., & Sheikh, Y. (2017). Realtime multi-person 2D pose estimation using part affinity fields. *Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition* (CVPR), 21–26 July 2017, Honolulu, HI, USA (pp. 1302–1310). https://doi.org/10.1109/cvpr.2017.143

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

Chen, Y., Zhou, J., Gao, J., Gao, G., Wang, S., & Zhang, W. (2021). Joint gaze estimation and facial expression for student engagement prediction in collaborative learning. *2021 IEEE International Conference on Engineering, Technology & Education* (TALE), 5–8 December 2021, Wuhan, China (pp. 703–707). IEEE. https://doi.org/10.1109/tale52509.2021.9678844

Clark, H. H., & Brennan S. E. (1991). Grounding in communication. In L. Resnick, J. Levine, & S. Teasley (Eds.), *Perspectives on socially shared cognition* (pp. 127–149). American Psychological Association. https://doi.org/10.1037/10096-006

D'Angelo, S., & Schneider, B. (2021). Shared gaze visualizations in collaborative interactions: Past, present and future. *Interacting With Computers*, *33*(2), 115–133. https://doi.org/10.1093/iwcomp/iwab015

D'Angelo, S., Brewer, J., & Gergle, D. (2019). Iris: A tool for designing contextually relevant gaze visualizations. *Proceedings of the 11th ACM Symposium on Eye Tracking Research & Applications* (ETRA '19), 25–28 June 2019, Denver, CO, USA (Article 79). https://doi.org/10.1145/3314111.3318228

D'Angelo, S., & Gergle, D. (2016). Gazed and confused: Understanding and designing shared gaze for remote collaboration. *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems* (CHI '16), 7–12 May 2016, San Jose, CA, USA (pp. 2492–2496). ACM Press. https://doi.org/10.1145/2858036.2858499

Denisova, O. A., Lekhanova, O. L., & Gudina, T. V. (2020). Problems of distance learning for students with disabilities in a pandemic. *SHS Web of Conferences*, *87*, 00044. https://doi.org/10.1051/shsconf/20208700044

Djamasbi, S., Siegel, M., & Tullis, T. S. (2012). Faces and viewing behavior: An exploratory investigation. *AIS Transactions on Human–Computer Interaction*, *4*(3), 190–211. https://doi.org/10.17705/1thci.00046

Efron, B., & Tibshirani, R. J. (1993). *An introduction to the bootstrap*. Chapman & Hall/CRC. https://doi.org/10.1201/9780429246593

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

Garbarino, M., Lai, M., Bender, D., Picard, R. W., & Tognetti, S. (2014). Empatica E3: A wearable wireless multi-sensor device for real-time computerized biofeedback and data acquisition. *Proceedings of the 4th International Conference on Wireless Mobile Communication and Healthcare: Transforming Healthcare Through Innovations in Mobile and Wireless Technologies* (MOBIHEALTH 2014), 3–5 November 2014, Athens, Greece (pp. 39–42). IEEE. https://doi.org/10.4108/icst.mobihealth.2014.257418

Guo, Z., & Barmaki, R. (2020). Deep neural networks for collaborative learning analytics: Evaluating team collaborations using student gaze point prediction. *Australasian Journal of Educational Technology*, *36*(6), 53–71. https://doi.org/10.14742/ajet.6436

Hartshorn, K. J., & McMurry, B. L. (2020). The effects of the COVID-19 pandemic on ESL learners and TESOL practitioners in the United States. *International Journal of TESOL Studies*, *2*(2), 140–156. https://doi.org/10.46451/ijts.2020.09.11

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

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. https://doi.org/10.1016/j.learninstruc.2012.11.004

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

Kizilcec, R. F., Papadopoulos, K., & Sritanyaratana, L. (2014). Showing face in video instruction: Effects on information retention, visual attention, and affect. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (CHI '14), 26 April–1 May 2014, Toronto, ON, Canada (pp. 2095–2102). ACM Press. https://doi.org/10.1145/2556288.2557207

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

Li, W., Wang, F., Mayer, R. E., & Liu, H. (2019). Getting the point: Which kinds of gestures by pedagogical agents improve multimedia learning? *Journal of Educational Psychology*, *111*(8), 1382–1395. https://doi.org/10.1037/edu0000352

McAlpin, E., Levine, M., & Plass, J. L. (2023). Comparing two whole task patient simulations for two different dental education topics. *Learning and Instruction*, *83*, 101690. https://doi.org/10.1016/j.learninstruc.2022.101690

Mason, L., Pluchino, P., & Tornatora, M. C. (2015). Eye-movement modeling of integrative reading of an illustrated text: Effects on processing and learning. *Contemporary Educational Psychology*, *41*, 172–187. https://doi.org/10.1016/j.cedpsych.2015.01.004

Mautone P. D., & Mayer R. E. (2001). Signaling as a cognitive guide in multimedia learning. *Journal of Educational Psychology*, *93*(2), 377–389. https://doi.org/10.1037//0022-0663.93.2.377

Mayer, R. E. (2005). Cognitive theory of multimedia learning. In R. Mayer (Ed.), *The Cambridge handbook of multimedia learning* (pp. 31–48). Cambridge University Press. https://doi.org/10.1017/cbo9780511816819.004

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

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

Melnyk, R., Campbell, T., Holler, T., Cameron, K., Saba, P., Witthaus, M. W., Joseph, J., & Ghazi, A. (2021). See like an expert: Gaze-augmented training enhances skill acquisition in a virtual reality robotic suturing task. *Journal of Endourology*, *35*(3), 376–382. https://doi.org/10.1089/end.2020.0445

Mundy, P., Sigman, M., & Kasari, C. (1990). A longitudinal study of joint attention and language development in autistic children. *Journal of Autism and Developmental Disorders*, *20*(1), 115–128. https://doi.org/10.1007/bf02206861

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

Paciej-Woodruff, A. (2021). The case for your face: Teacher presence in asynchronous education courses. Stop feeling uncomfortable and start recording your face to humanize the online experience. *Society for Information Technology & Teacher Education International Conference* (SITE), 29 March 2021, Online (pp. 535–539). Association for the Advancement of Computing in Education (AACE). https://www.learntechlib.org/primary/p/219181/

Papoutsaki, A., Sangkloy, P., Laskey, J., Daskalova, N., Huang, J., & Hays, J. (2016). WebGazer: Scalable webcam eye tracking using user interactions. In S. Kambhampati (Ed.), *Proceedings of the 25th International Joint Conference on Artificial Intelligence* (IJCAI '16), 9–15 July 2016, New York, NY, USA (pp. 3839–3845). AAAI Press/International Joint Conferences on Artificial Intelligence. https://www.ijcai.org/Proceedings/16/Papers/540.pdf

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

Sauter, M., Wagner, T., & Huckauf, A. (2022). Distance between gaze and laser pointer predicts performance in video-based e-learning independent of the presence of an on-screen instructor. *2022 Symposium on Eye Tracking Research and Applications* (ETRA '22), 8–11 June 2022, Seattle, WA, USA (Article 26). ACM Press. https://doi.org/10.1145/3517031.3529620

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), 010502. https://doi.org/10.11613/bm.2021.010502

Schneider, B. (2020). A methodology for capturing joint visual attention using mobile eye-trackers. *JoVE,* (155), e60670. https://doi.org/10.3791/60670-v

Schneider, B. (2023). Three Challenges in Implementing Multimodal Learning Analytics in Real-World Learning Environments. *Learning: Research and Practice*, 1-10. https://doi.org/10.1080/23735082.2023.2270611

Schneider, B., & Bryant, T. (2024). Using mobile *dual* eye-tracking to capture cycles of collaboration and cooperation in co-located dyads. *Cognition and Instruction*, *42*(1), 26–55. https://doi.org/10.1080/07370008.2022.2157418

Schneider, B., Davis, R., Martinez-Maldonado, R., Biswas, G., Worsley, M., & Rummel, N. (2024). Stepping outside the ivory tower: How can we implement multimodal learning analytics in ecological settings, and turn complex temporal data sources into actionable insights? *Proceedings of the 17th International Conference on Computer-Supported Collaborative Learning* (CSCL 2024), 10–14 June 2024, Buffalo, NY, USA (pp. 323–330). International Society of the Learning Sciences. https://doi.org/10.22318/cscl2024.259119

Schneider, B., Dich, Y., & Radu, I. (2020). Unpacking the relationship between existing and new measures of physiological synchrony and collaborative learning: A mixed methods study. *International Journal of Computer-Supported Collaborative Learning*, *15*(1), 89–113. https://doi.org/10.1007/s11412-020-09318-2

Schneider, B., Feng, D., & Sung, G. (2023). Joint visual attention predicts learning in 1-on-1 remote teaching: A dual eye-tracking study. *Proceedings of the 16th International Conference on Computer-Supported Collaborative Learning* (CSCL 2023), 10–15 June 2023, Montréal, QC, Canada (pp. 83–90). International Society of the Learning Sciences. https://doi.org/10.22318/cscl2023.849100

Schneider, B., & Pea, R. (2013). Real-time mutual gaze perception enhances collaborative learning and collaboration quality. *International Journal of Computer-Supported Collaborative Learning*, *8*(4), 375–397. https://doi.org/10.1007/s11412-013-9181-4

Schneider, B., Sharma, K., Cuendet, S., Zufferey, G., Dillenbourg, P., & Pea, R. (2018). Leveraging mobile eye-trackers to capture joint visual attention in co-located collaborative learning groups. *International Journal of Computer-Supported Collaborative Learning*, *13*(3), 241–261. https://doi.org/10.1007/s11412-018-9281-2

Schneider, B., Sung, G., Chng, E., & Yang, S. (2022). How Can High-Frequency Sensors Capture Collaboration? A Review of the Empirical Links Between Multimodal Metrics and Collaborative Constructs. *Sensors*, *21*(24), 8185.

Sharma, K., Alavi, H. S., Jermann, P., & Dillenbourg, P. (2016). A gaze-based learning analytics model: In-video visual feedback to improve learner's attention in MOOCs. *Proceedings of the 6th International Conference on Learning Analytics and Knowledge* (LAK '16), 25–29 April 2016, Edinburgh, UK (pp. 417–421). ACM Press. https://doi.org/10.1145/2883851.2883902

Sharma, K., D'Angelo, S., Gergle, D., & Dillenbourg, P. (2016). Visual augmentation of deictic gestures in MOOC videos. In C. K. Looi, J. L. Polman, U. Cress, & P. Reimann (Eds.), *Transforming learning, empowering learners: The international conference of the learning sciences (ICLS) 2016* (Vol. 1, pp. 202–209). International Society of the Learning Sciences. https://repository.isls.org/handle/1/117

Sharma, K., Jermann, P., & Dillenbourg, P. (2014). "With-me-ness": A gaze-measure for students' attention in MOOCs. *Learning and Becoming in Practice: Proceedings of the International Conference of the Learning Sciences* (ICLS '14), 23–27 June 2014, Boulder, CO, USA (Vol. 2, pp. 1017–1022). International Society of the Learning Sciences. https://repository.isls.org//handle/1/924

Sharma, K., Jermann, P., & Dillenbourg, P. (2015). Displaying teacher's gaze in a MOOC: Effects on students' video navigation patterns. In G. Conole, T. Klobučar, C. Rensing, J. Konert, & E. Lavoué (Eds.), *Design for teaching and learning in a networked world: 10th European conference on technology enhanced learning, EC-TEL 2015, Toledo, Spain, September 15–18, 2015, proceedings* (pp. 325–338). Springer Cham. https://doi.org/10.1007/978-3-319-24258-3_24

Špakov, O., Niehorster, D., Istance, H., Räihä, K.-J., & Siirtola, H. (2019). Two-way gaze sharing in remote teaching. In D. Lamas, F. Loizides, L. Nacke, H. Petrie, M. Winckler, & P. Zaphiris (Eds.), *Human–computer interaction – INTERACT 2019: 17th IFIP TC 13 international conference, Paphos, Cyprus, September 2–6, 2019, proceedings, part II* (pp. 242–251). Springer Cham. https://doi.org/10.1007/978-3-030-29384-0_16

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

Sung, G., Feng, T., & Schneider, B. (2021). Learners learn more and instructors track better with real-time gaze sharing. *Proceedings of the ACM on Human–Computer Interaction* (CSCW1, Vol. 5, Article 134). ACM Press. https://doi.org/10.1145/3449208

Tomasello, M. (1995). Joint attention as social cognition. In C. Moore & P. J. Dunham (Eds.), *Joint attention: Its origins and role in development* (pp. 103–130). Lawrence Erlbaum. https://doi.org/10.4324/9781315806617

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

Westfall, J. (2015, May 26). Think about total N, not n per cell. *Cookie Scientist.* https://web.archive.org/web/20190216111000/http://jakewestfall.org/blog/index.php/2015/05/26/think-about-total-n-not-n-per-cell/#expand

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

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. https://doi.org/10.1111/jcal.12568