Showing Face in Video Instruction: Effects on Information Retention, Visual Attention, and Affect

René F. Kizilcec[1], Kathryn Papadopoulos[2,3], Lalida Sritanyaratana[2]

Department of Communication[1]
Stanford University
kizilcec@stanford.edu

HCI Group[2]
Stanford University
lalida@google.com

Customer Experience[3]
Citrix
kathryn.papadopoulos@citrix.com

ABSTRACT
The amount of online educational content is rapidly increasing, particularly in the form of video lectures. The goal is to design video instruction to facilitate an experience that maximizes learning and satisfaction. A widely used but understudied design element in video instruction is the overlay of a small video of the instructor over lecture slides. We conducted an experiment with eye-tracking and recall tests to investigate how adding the instructor's face to video instruction affects information retention, visual attention, and affect. Participants strongly preferred instruction with the face and perceived it as more educational. They spent about 41% of time looking at the face and switched between the face and slide every 3.7 seconds. Consistent with prior work, no significant difference in short- and medium-term recall ability was found. Including the face in video instruction is encouraged based on learners' positive affective response. More fine-grained analytics combining eye-tracking with detailed learning assessment could shed light on the mechanisms by which the face aids or hinders learning.

Author Keywords
Audiovisual Instruction; Multimedia Learning; Eye-tracking.

ACM Classification Keywords
K.3.1. Computers and Education: Computer Uses in Education, Distance learning.

INTRODUCTION
Do you remember those educational TV programs your science teachers showed in class? Audiovisual instruction is not a novelty of the 21st Century; it has been around for many decades, whether in classrooms, on TV, or for corporate training. However, the increasing ubiquity of broadband Internet access has fueled the rapid proliferation of audiovisual educational material in the last few years. In 2012, 32% of higher education students took at least one course online [1]. And with the recent development of online asynchronous computer-mediated communication platforms for educational content, access to education, mostly in the form of audiovisual instruction, continues to extend far beyond traditional academic institutions.

One of the largest providers of massive open online courses (MOOCs), Coursera, went from zero to 2.9 million registered users from more than 220 countries in their first year of operation [38]. The majority of courses offered by Coursera and other online education providers, including corporate training providers, are built around audiovisual instruction. Hundreds of hours are spent preparing, recording, and editing original video lectures for a typical online course [20]. Research on how to effectively design audiovisual instruction undoubtedly has the potential to improve the educational experience of millions of people today.

A major concern with online courses has been the sense of isolation that many learners report and that can hinder their ability to learn [29]. Recent ethnographic findings suggest that video-based courses facilitate a richer interaction between learners and the instructor, who reportedly seems more present and real to learners; moreover, learners in such courses find the learning experience similar to face-to-face instruction [4]. The influence of instructional media, such as video lectures, on learning has been the topic of the "media effects" debate between Clark, who holds that the choice of medium has no influence on learning outcomes [7, 8], and Kozma, who argues that carefully matching media with learners and instructional goals can lead to more effective learning experiences [21, 22]. Kozma emphasizes the importance of attributes of various media over the specific media themselves; for instance, the extent to which social cues can be embedded in a medium.

A number of studies in multimedia learning research have investigated multiple channel presentation and the resulting phenomenon of split attention [13, 27, 34]. Similarly, research on instruction in physical classrooms has explored the attention division between the professor and blackboard or slides [33]. Following this line of research and observing that an increasing number of online instruction is presented as full-screen lecture slides with a small embedded video overlay of the instructor speaking (Figure 1), we investigate the effects of the instructor's face in video instruction on visual attention, information retention, and learner affect.

BACKGROUND

Cognitive Processing, Learning, and Affect

CHI 2014, April 26–May 1, 2014, Toronto, Ontario, Canada.

ACM 978-1-4503-2473-1/14/04..$15.00.
http://dx.doi.org/10.1145/2556288.2557207

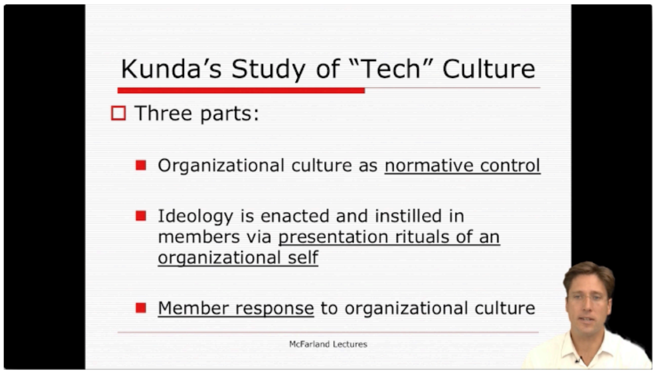

Figure 1. Example of video instruction with lecture slides in the background and a picture-in-picture of the instructor's face.

Meaningful learning occurs when cognitive processing of instructional material is moderate: too little processing results in subpar learning, while too much processing can lead to cognitive overload, potentially inhibiting learning [28]. We can use Baddeley's Theory of Working Memory [2] as a lens to understand cognitive processes in multimedia learning. It posits that separate processing units are employed for different input modalities: one area of working memory, called the visual-spatial 'sketchpad', stores visual input while another area, the phonological loop, stores auditory information. According to this model, the video of the instructor's face would compete for visual-spatial cognitive resources with the slide content while the instructor's narration is processed separately though potentially supported by non-verbal information encoded in the instructor's face (e.g., gestures and facial expressions).

Clark and Mayer [6] present a series of empirically established principles to reduce cognitive load in the design of multimedia instruction. For instance, if the content of the presentation is overloading the visual channel, some of the information should be moved to the auditory channel for processing. Extraneous material should be removed, as it may interfere with the resource allocation for processing the important learning materials. And instructional design that induces switching between materials that are presented through different modalities (text, pictures, etc.) may increase cognitive load. The design of much audiovisual instruction has been informed by such insights from multimedia learning theory. Prior research in the multimedia learning tradition on the instructor's face has investigated its effect on learning (perceived, and recall and transfer ability), perceived social presence, cognitive load, and self-reported affective measures (e.g., usefulness and comfort) [15, 24, 25]. The present study is the first that combines information recall and affective measures with information on how learners watch a video lecture with and without the instructor's face.

At first sight, the instructor's face might be considered 'extraneous', as all lecture-relevant information is encoded on the slides and in the narration. As a result, the additional extraneous processing could hinder cognitive processing of important information and ultimately hinder learning. A more positive view on the face is put forward by Clark and Mayer [6] who emphasize the effectiveness of social cues from the instructor, because they trigger social responses in the learner [32] and encourage deeper engagement with the lecture content. Wang et al. [39] report higher learning outcomes as a result of learners responding with politeness to a pedagogical agent, because they treat it as a social actor. Moreover, learners who worked with pedagogical agents with a human voice and that exhibit more natural, human-like gestures, facial expression, and eye gaze performed better on knowledge transfer tests than learners working with a less humanoid agent [26].

Experimental evidence on the effect of including the instructor's face in lecture videos, however, does not support the predictions of this social argument. Mayer's [25] image principle, which summarizes multiple studies' insignificant findings for learning outcomes (e.g. [30]), states that adding a picture or video of the instructor to multimedia instruction does not necessarily support learning. The image principle has also been confirmed in more recent studies. Homer, Plass, and Blake [15] found no significant difference in recall and transfer knowledge as a result of adding a small video of the speaker to a video of slides from a conference talk with narration. Nevertheless, learners who saw the speaker's face did not report a greater sense of social presence than those who did not see the speaker, which might suggest that the social cue was too weak to induce positive social responses in learners. In light of the conflicting views on the effect of the instructor's face on learning outcomes, we pose the following research question:
RQ1: Are recall scores higher, lower, or equal if the instructor's face is present in the lecture video than when it is absent?

Beyond learning outcomes, there is also competing empirical evidence on learners' affective response to the instructor's face in video lectures. Consistent with predictions from the "media equation" [32], greater perceived social presence in online learning environments has repeatedly been associated with increased learner satisfaction and perceived learning (see [9] for a review). However, in a recent investigation into affective responses to the instructor's face, learners with low technological efficacy reported lower perceived learning, social presence, and video usefulness when the face was present [24]. Moreover, Homer et al. [15] report higher levels of cognitive load for learners who saw the video of the speaker; too high cognitive load is bound to damage the learning experience. Given the conflicting evidence on learner's affective response to the face, we pose the following research question:
RQ2: Do learners prefer audiovisual instruction with or without the instructor's face? How do learners rate the experience with and without the instructor's image on a set of descriptive adjectives?

In a large literature on individual differences in learning, preferences for the modality of information presentation (written, spoken, visual aids, etc.) have been studied extensively. While individuals may choose representations that align with their preferences, the literature is mixed on whether or not improvement in actual learning occurs [10]. For instance, when a foreign-language story was supplemented with visual aids

(e.g., relevant pictures), students with a visual preference understood it better, while those with a verbal preference understood it better with a written translation of the text [31]. Although Homer et al. [15] found no effect of information preference on learning outcomes, they did find the presence of the speaker's face to affect viewers' cognitive load. Specifically, those who preferred visual information reported higher cognitive load without the speaker's video, while those who preferred verbal information experienced higher load with the video. A higher level of cognitive load indicates additional cognitive processing that could either support learning (germane cognitive load) or hinder learning (extraneous cognitive load) [36]. Based on this evidence for individual differences in preferences for information presentation, we formulate the following hypothesis:
H1: Preferences for information presentation moderates the effect of the presence of the face on recall ability.

Visual Attention and Faces

Eye-tracking generates a wealth of information but it can be challenging to identify what information is processed and retained based on people's gaze. The eye-mind hypothesis [18] suggests that there exists a link between human gaze and attention, positing that people process the information that they visually attend to.

Historically, the transfer of knowledge among humans has been mediated primarily through face-to-face communication. Preferential looking studies have confirmed people's innate tendency to attend to faces over, for example, scrambled faces or blank stimuli [12]. The human attraction to faces has been shown to already develop in infants who attend favorably to faces and face-like configurations [17]. From early on, people are attracted to stimuli that promote social interaction and communication [11]. Consequently, we would expect showing the instructor's face in video lectures to have a profound impact on learners' allocation of attention. We therefore pose the following descriptive question: What proportion of time do learners spend looking at the face when it is present? And for how long do they dwell on the face and the slides?

Eye-tracking research on multimedia learning is only in its beginnings, but strongly encouraged by previous work to gain insights into what learners attend to, for how long, and in which order [37]. In a line of eye-tracking research on the spatial contiguity principle, which asserts that placing words near corresponding graphics promotes learning, transitions between words and graphics were interpreted as evidence for integrative cognitive processing, which occurs when information from the visual and auditory channels in working memory are integrated [16, 34, 14]. Schmidt-Weigand et al. [34] found that learners transition around 0.22 times per second between a graphic and corresponding written text in an animation on lightning formation. Similarly, the instructor's face competes for visual attention with content presented on the lecture slides, which induces split visual attention. Although the text shown on the slides does not describe the face, the instructor who is represented by the face tends to refer to information on the slides in speech and gesture. This prompts the learner to engage in similar eye movements to integrate information from the two sources, though it is unclear how this behavior compares to the spatial contiguity case. Thus, we pose the following descriptive question: How frequently do learners transition between fixating on the face and the slide content when the face is present?

Finally, in the case of split visual attention, where learners transition between competing visual stimuli, the tension, and thus the transition rate, that develops between the stimuli is likely higher in learners who prefer learning from visual than verbal information. Accordingly, we hypothesize:
H2: The transition rate is higher for learners who prefer visual than verbal information.

METHODS

Participants

Undergraduate and graduate students (n=22; 11 female; aged 18 to 24) were recruited from the participant pool of a major U.S. university. It was ensured that none of the participants were knowledgeable on the topic of the video lecture. Participants were awarded 1 unit of course credit upon completion of the study.

Experimental Design

This study used a 2 (face, no face) x 7 (repetitions) repeated-measures design. Participants were randomly assigned to watch a video lecture in which a video of the instructor is either present or absent in 3 or 4 of the 7 video segments. This yielded 4 segments in which the face was present for a random half of the participants but absent for the other half, and 3 segments for which it was reversed. To estimate the average causal effect of the instructor's face on information recall ability, we compute the difference in recall test scores between participants who watched lecture segments with the face and those who watched them without.

Materials

Stimulus

The stimulus was presented as a video with audio. The video starts with a 30 second resting period (black screen with white text "Resting Period") followed by 14 minutes 18 seconds of continuous instruction on a topic in organizational Sociology by Professor Dan McFarland. In the instruction period the instructor's face alternates between being present and absent approximately every 2 minutes (coinciding with slide changes), yielding a total of seven segments (four with face in one order, three with face in the other order). When the face was present, participants viewed a lecture video with a picture-in-picture (PiP) video of the instructor's face in the lower right corner. When it was absent, there was no PiP of the instructor's face in the lecture video, only lecture slides.

Eye tracker

Eye position was measured in x-y coordinates of the display monitor using the SensoMotoric Instruments (SMI) iView X RED tracker. The device was mounted to the bottom of the computer monitor on which the lecture video was displayed. It operates at a distance of 60 - 80cm and has a high accuracy

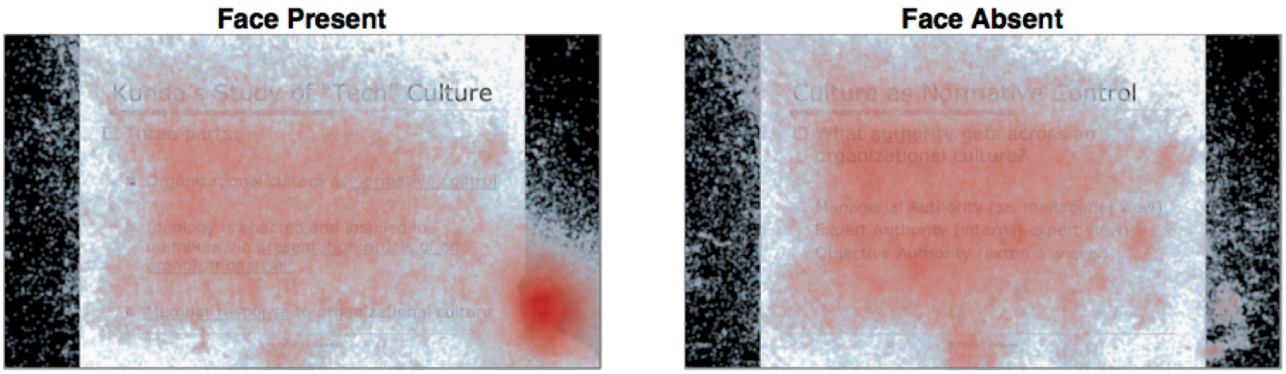

Figure 2. Heatmap of visual attention aggregated for all segments where the face was present or absent with exemplary screenshot in the background.

of 0.4 degrees. Its contact-free setup allows for free head movement of 40cm x 20cm at 70cm distance.

Questionnaires

The pre-stimulus questionnaire contained questions on gender, age, amount of experience with video lectures, prior subject knowledge, and preference for information presentation in educational contexts (two sentence-completion items from the index of learning styles questionnaire [35]: "I prefer to get new information in i. pictures, diagrams, graphs, or maps, ii. written directions"; "I like teachers who i. put a lot of diagrams on the board, ii. spend a lot of time explaining.").

The post-stimulus questionnaire contained questions on participants' affective response to lecture segments with and without the instructor's face. Participants were asked to rate what parts of the lecture video they liked better: the parts where the instructor's face was present or absent on a 10-point bipolar scale. Participants also rated how well each of a list of adjectives describes their lecture experience when the face was present and separately for when the face was absent. Adjectives included 'helpful', 'useful', 'frustrating', 'annoying', 'confusing' and were rated on a 10-point bipolar scale from "describes very poorly" to "describes very well".

Knowledge Recall Tests

Two tests were administered to measure participants' ability to recall information from the lecture. The short-term recall test was administered following the post-stimulus questionnaire about three minutes after watching the lecture. A medium-term recall test was administered five days after watching the lecture. Both recall tests contained the same 19 questions presented in random order, consisting of seven multiple choice (MC) and twelve select-all-that-apply items. In order to reduce variance from guessing, participants were instructed to select the "I don't know" (IDK) answer option if they did not know the correct answer.

Test scores were computed by awarding one point for each correct MC answer and one point for each correctly selected or unselected choice option in select-all-that-apply questions, zero otherwise. Questions that were answered IDK also received zero points. All scores were normalized by the total number of possible points.

Procedure

After granting consent, the experimenter escorted the participant to the eye-tracking room, which contained a chair facing a large computer monitor. The participant completed the pre-test on a laptop in the absence of the experimenter. The experimenter returned to remove the laptop and instructed the participant to pay attention to the lecture as there would be a short quiz on the material afterwards. The experimenter then turned off the lights in the room, exited, and checked the participant's eye position in the experimenter room. The participant was instructed via radio to alter the chair position until their eye level was centered to the monitor. Before the video stimulus started, participants were asked to follow a dot on the screen with their eyes for calibration. Calibration was repeated until satisfactory accuracy levels were achieved.

The participant watched the stimulus without pauses and was not allowed to take notes. Once the lecture was over, the experimenter turned on the lights and handed the participant a laptop to complete the post-test. After completion, the participant was asked to guess the purpose of the study and was then debriefed on the experiment. Most participants thought the study was conducted to inform video lecture design and mentioned the instructor's face. Five days later, the participant received an email with a link to the medium-term recall test to complete on the same day.

RESULTS

Affective Response

All participants' ratings of which video segments they preferred were strongly in favor of showing the face, with 15 out of 22 participants extremely preferring segments with the face. This clearly addresses the first part of research question RQ2. All eight participants who prefer learning from visual information extremely preferred the face, while responses from those who prefer learning from verbal information were more varied but still positive.

To answer the second part of RQ2, a factor analysis on participants' ratings of how well each adjective applies to their experience was performed, yielding two factors: one on educational value (helpful, useful) and the other on learning experience (frustrating, annoying, confusing). Participants indicated that the sections of the lecture video where the face was present were more helpful and useful, $t(21)$=5.58, p<0.001, Cohen's D=1.19, while the sections without the face were more frustrating, annoying, and confusing, $t(21)$=4.36, p<0.001, D=0.93. Furthermore, participants who self-reported to prefer learning from verbal rather than visual information more strongly indicated that sections with the face were more helpful and useful, $t(18)$=2.19, p=0.04, D=0.9.

Visual Attention

In this section we present insights from the eye-tracking data on how participants watched the lecture video. Figure 2 shows a heatmap of where participants looked during segments when the face was present and when it was absent. An example of a lecture slide with and without the face is shown in the background to provide context. It stands out that the face receives considerable visual attention when it is present. In fact, when the face was present, participants spent 41% of time looking at it on average (median=39%, SD=14.5%, ranging from 9% to 60%).

Visual attention is typically measured in the form of fixations, which describe durations of around 500 milliseconds that a person spends looking at a small area on the screen. At times when the instructor's face is present, visual attention is split between the face and the lecture slide content. Table 1 summarizes the average number of fixations, the average fixation length, and average median fixation length (in seconds). While the number of fixations does not differ significantly, participants dwelled for longer periods on the face than on the lecture slide content, even though the area of the face is smaller than the slide.

It is worth noting that participants' self-reported preference for information presentation was not significantly associated with their lecture watching behavior. In particular, participants who reported to prefer learning from written materials did not spend more time looking at the slide content than those who reported to prefer visual materials, $mean_{\text{visual}}$=40.6%, SD_{visual}=15.1%, $mean_{\text{written}}$=42.7%, SD_{written}=13.9%. To investigate how frequently learners transition between fixating on the face and the slide content when the face is present, we computed the transition rate (transitions per second) following Schmidt-Weigand et al. [34] (using the previous definition of a fixation as lasting at least 500 ms). We find that participants transition 0.27 times per second on average (SD=0.09, median=0.28, ranging from 0.11 to 0.49) which is considerably higher than transition rates reported by Schmidt-Weigand et al. [34]. Moreover, we hypothesized that a preference for visual versus verbal information would induce higher transition rates (H2). The data supports this hypothesis with visual-preference learners switching 1.3 times more frequently than those with a verbal preference, $t(20)$=2.0, p=0.060, D=0.88.

Table 1. Mean (and SD) of fixation count and length while the face is present.

Target	Count	Length	Median Length
Instructor Face	57.0 (27.0)	11.5s (5.4)*	2.9s (2.0)*
Lecture Slide	58.4 (23.0)	2.7s (1.4)*	1.7s (0.8)*

* significantly different at p<0.05

Information Retention

Scores on the short-term recall test were approximately normally distributed with mean 0.588 and standard deviation 0.163, ranging from 0.273 to 0.832. Scores on the medium-term recall test were left skewed (probably due to question familiarity) and distributed with $mean$=0.532, SD=0.233, ranging from 0.105 to 0.857. Average test scores were not significantly different at the 5% level between test and retest, $t(21)$=1.87, p=0.075, D=0.40, but there appears to be a downward trend in scores.

To investigate the research question of whether the instructor's face affects recall (RQ1), each answer option was coded by whether the face was present or absent when the relevant information was presented in the lecture. As participants were randomly assigned to a stimulus order, half the participants answered questions for which the face was present, while the other half answered the same questions but did not see the face. As a result, there is no issue of question-level confounding for this comparison.

Two repeated-measures MANOVAs were performed to simultaneously test short- and medium-term recall scores repeated over stimulus segments. The first MANOVA compares scores for odd segments (1, 3, 5, 7) in which participants assigned to stimulus order 1 saw the face and the rest did not. The second test is performed for even segments (2, 4, 6) in which participants assigned to order 2 saw the face and the rest did not. Both tests yielded a significant repetition effect (i.e. scores changed over segments), but no significant main effect or interaction (with the segment number) associated with the presence of the face, $F(2,19)$=0.275, p=0.76, 95% CI=[-0.193, 0.322] for odd segments, and $F(2,19)$=0.709, p=0.50, 95% CI=[-0.759, 0.211] for even segments. A power analysis was performed to compute the required effect size that could be detected with 95% confidence given 80% power, with 22 participants and four repeated measures, and a 0.2 correlation between repeated measures: a medium to large effect size f of 0.4 is detectable under these circumstances.

Figure 3 illustrates test scores for each segment with 95% confidence intervals (clipped to [0-1]) by whether the participant watched the segment with or without the instructor's face. Notably, confidence intervals between groups overlap considerably for all segments. This provides strong evidence that whether or not the face is present has no effect on recall ability, neither immediately after the lecture, nor five days later (medium-term recall). Moreover, we find no evidence in support of hypothesis H1 that learning preferences moderate the effect of the face on learning.

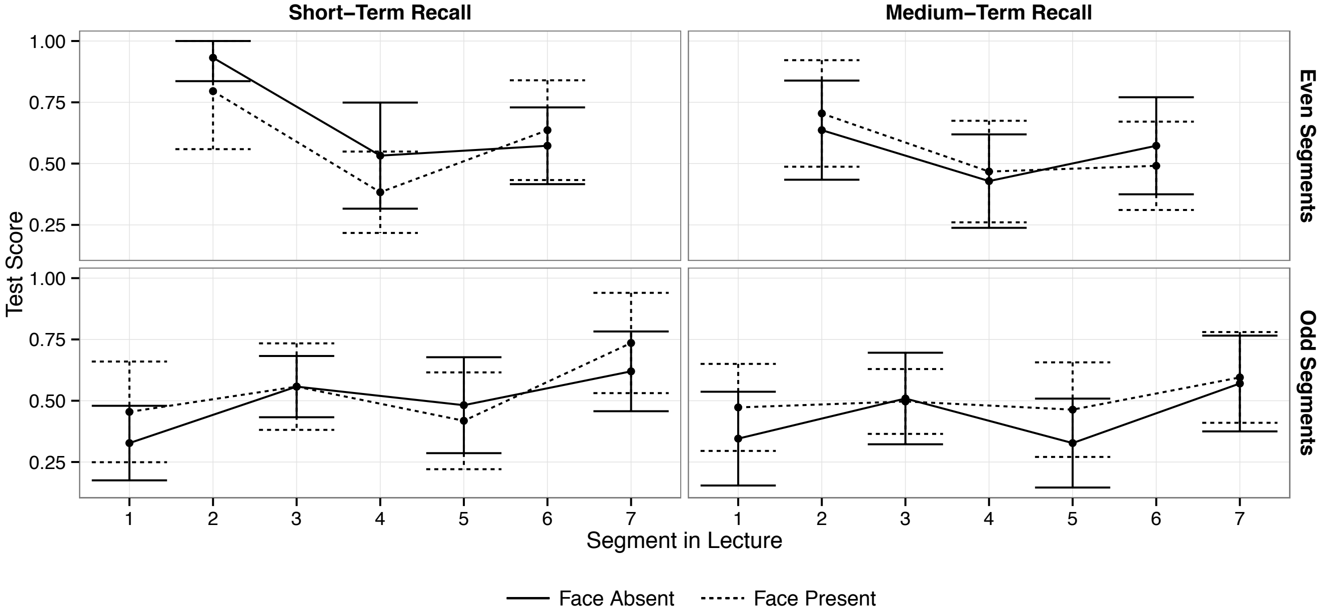

Figure 3. Average test scores with 95% confidence intervals by recall type for each segment comparing scores with the face present and absent.

DISCUSSION

The present study suggests that although learners strongly prefer video instruction with the instructor's face and their watching behavior is profoundly changed by including the face, they do not perform significantly better or worse on knowledge recall tests compared to without the face. This finding simultaneously reflects the established null results for learning outcomes as summarized in the image principle [25] as well as established positive affective responses to social cues in learning environments [9]. This positive affective response to the face is likely to boost learner motivation, which positively affects the amount of time learners are prepared to devote to learning [5]. Beyond learning and affect, this investigation is the first to uncover how the presence of the face changes learners' viewing behavior: on average, learners spent 41% of time looking at the face when it was present and transitioned between looking at the face and slides every 3.7 seconds. Despite being a relatively coarse characterization of learners' watching behavior, it emphasizes the great impact on visual attention of including the instructor's face. These findings suggest that a more fine-grained analysis of individual learners' watching patterns combined with detailed learning assessments could provide very rich insights into how the instructor's face might aid or hinder learning.

The comparison of individual differences in learning preference (learning from visual versus verbal information) produced two insights. First, seeing the face was rated as more useful by those who prefer learning from verbal information, which is not surprising, as although the face is visual (an image), it conveys non-verbal information which supports the processing of verbal information, thus alleviating the amount of cognitive load. According to this view, the face is largely processed in the verbal channel, because it shows the instructor speaking, which encourages lipreading and provides other non-verbal information. Second, as hypothesized, learners who prefer learning from visual information transitioned more frequently (by a factor of 1.3) between the face and slides than those with a verbal preference. The transition rate can be a measure of the amount of integrative processing learners engage in, though a high transition rate could also signify difficulties integrating the information [14]. Learners probably switch between the face and the slides to compare the instructor's narration to what is illustrated or written on the slides.

Our findings speak in favor of presenting the instructor's face in video instruction from a learner affect perspective, but are subject to certain limitations. Our study participants–undergraduate and graduate students–represent an appropriate sample for a study on lecture videos in a brick-and-mortar university setting, but are less representative of the population of lifelong learners and corporate trainees who rely heavily on video instruction. Moreover, the study was conducted in a highly controlled environment, which might have induced demand characteristics that affected participants' performance and watching behavior. Specifically, participants were told they would be tested on the material in the lecture video, which may have led to increased apprehension and desire to be a 'good participant'. Participants were told that their eye movements would be recorded, which may have induced unnatural lecture-watching behavior. Mayer and Moreno [28] suggest that if both visual and auditory channels are overloaded, segmenting (adding breaks) may reduce cognitive load. In our study, the lecture video was shown without the opportunity to pause, rewind, or take notes, which might have induced higher levels of cognitive load than would have been experienced without such constraints. Although an authentic video lecture from a MOOC was employed as a stimulus in this study, the generalizability of our findings is constrained by significant variation in video length and other attributes of audiovisual instruction between online courses.

Our study observed a specific type of interaction with video instruction, namely one where learners attentively follow the lecture content because they expect to be tested afterwards but are not taking lecture notes. Historically, this type of engagement has been very common, but online learning opportunities, such as MOOCs, have created space for less performance-oriented types of engagement, for example auditing or exploring a course [19]. A majority of MOOC learners do not take the course with the intention of performing well on tests, which we might expect to be reflected in a less conscious effort to remember lecture content, more frequent pausing, and casual note-taking, if any. The variety in how learners interact with audiovisual instruction poses a challenge for the design of video instruction to facilitate personalized learning experiences.

Future work could attempt to distinguish between patterns of lecture watching and how instructional design affects learners differentially depending on their objectives. This line of research could shed light onto how video instruction can be customized to learning objectives. The present study employed knowledge tests for short- and medium-term recall memory as well as a measure of affect toward the lecture experience. Future research should use measures that go beyond information recall and assess deep understanding and conceptual knowledge in combination with eye-tracking measures to determine how including the instructor's face might aid or hinder learning. Finally, a longitudinal field experiment in which learners' interactions with video instruction and test performance is observed over time for thousands of learners in a MOOC could provide valuable insights on the effects of the instructor's face and other design elements in video instruction.

CONCLUSION

The unique contribution of this paper is to combine eye-tracking with affective and information recall measures to inform instructional design decisions by providing empirical evidence for the effects of a widely used but understudied design element: the instructor's face in video instruction. Our findings plausibly generalize to a group of potential learners within the usual constraints of the instructor's lecture style and the lecture topic presented in the video stimulus. A missing piece in the cost-benefit analysis for the instructional design decision on the instructor's face in video lectures is the financial and opportunity cost of filming and editing it. While further research into the long-term effects of including the instructor's face under more realistic circumstances are needed, this investigation could only uncover an affective benefit of the face.

Many institutions employ the picture-in-picture instructional format in their video lectures using expensive recording equipment and editing software [3]. Moreover, training instructors and teaching staff on its usage, as well as allocating time for editing, is a time-consuming addition to the course production process. For instance, the University of Pennsylvania estimates production costs for an online course at around $50,000, with videography as its top expense [23]. Considering that adding the instructor's face to lecture slides has no apparent effect on recall ability, further research into its effect on learning beyond recall is required to evaluate whether course producers should incur the cost of adding the face.

ACKNOWLEDGMENTS

We are grateful to Stanford's Office of the Vice Provost for Online Learning for supporting this research and Dan McFarland for allowing us to use his lecture video as our experimental stimulus. We would also like to thank Dean Eckles, Scott Klemmer, Clifford Nass, Andreas Paepcke, Candace Thille, and our CHI review committee for their helpful feedback.

REFERENCES

1. Allen, I. E., and Seaman, J. *Changing Course: Ten Years of Tracking Online Education in the United States.* ERIC, 2013.
2. Baddeley, A. Working memory. *Science 255*, 5044 (1992), 556–559.
3. Bixhorn, A. How moocs are evolving with video technology. *eCampus News* (September 2013).
4. Borup, J., West, R. E., and Graham, C. R. Improving online social presence through asynchronous video. *The Internet and Higher Education 15*, 3 (2012), 195–203.
5. Bransford, J. D., Brown, A. L., Cocking, R. R., et al. *How people learn*. National Academy Press Washington, DC, 2000.
6. Clark, R. C., and Mayer, R. E. *E-learning and the science of instruction: Proven guidelines for consumers and designers of multimedia learning*. Wiley. com, 2011.
7. Clark, R. E. Reconsidering research on learning from media. *Review of educational research 53*, 4 (1983), 445–459.
8. Clark, R. E. Media will never influence learning. *Educational technology research and development 42*, 2 (1994), 21–29.
9. Cui, G., Lockee, B., and Meng, C. Building modern online social presence: A review of social presence theory and its instructional design implications for future trends. *Education and Information Technologies* (2012), 1–25.
10. Dunn, R. S., and Dunn, K. J. *Teaching secondary students through their individual learning styles: Practical approaches for grades 7-12*. Allyn & Bacon, 1993.
11. Farroni, T., Johnson, M. H., Menon, E., Zulian, L., Faraguna, D., and Csibra, G. Newborns' preference for

face-relevant stimuli: Effects of contrast polarity. *Proceedings of the National Academy of Sciences of the United States of America 102*, 47 (2005), 17245–17250.

12. Goren, C. C., Sarty, M., and Wu, P. Y. Visual following and pattern discrimination of face-like stimuli by newborn infants. *Pediatrics 56*, 4 (1975), 544–549.
13. Hartman, F. R. Single and multiple channel communication: A review of research and a proposed model. *Audiovisual communication review 9*, 6 (1961), 235–262.
14. Holsanova, J., Holmberg, N., and Holmqvist, K. Reading information graphics: The role of spatial contiguity and dual attentional guidance. *Applied Cognitive Psychology 23*, 9 (2009), 1215–1226.
15. Homer, B. D., Plass, J. L., and Blake, L. The effects of video on cognitive load and social presence in multimedia-learning. *Computers in Human Behavior 24*, 3 (2008), 786–797.
16. Johnson, C. I., and Mayer, R. E. An eye movement analysis of the spatial contiguity effect in multimedia learning. *Journal of Experimental Psychology: Applied 18*, 2 (2012), 178.
17. Johnson, M. H., Dziurawiec, S., Ellis, H., and Morton, J. Newborns' preferential tracking of face-like stimuli and its subsequent decline. *Cognition 40*, 1 (1991), 1–19.
18. Just, M. A., and Carpenter, P. A. A theory of reading: From eye fixations to comprehension. *Psychological review 87* (1980), 329–354.
19. Kizilcec, R. F., Piech, C., and Schneider, E. Deconstructing disengagement: analyzing learner subpopulations in massive open online courses. In *Proceedings of the Third International Conference on Learning Analytics and Knowledge*, ACM (2013), 170–179.
20. Kolowich, S. The professors who make the moocs. *Chronicle of Higher Education 59*, 28 (2013), A20–A23.
21. Kozma, R. B. Learning with media. *Review of educational research 61*, 2 (1991), 179–211.
22. Kozma, R. B. Will media influence learning? reframing the debate. *Educational technology research and development 42*, 2 (1994), 7–19.
23. Lewin, T. Students rush to web classes, but profits may be much later. *The New York Times 6* (January 2013).
24. Lyons, A., Reysen, S., and Pierce, L. Video lecture format, student technological efficacy, and social presence in online courses. *Computers in Human Behavior 28*, 1 (2012), 181–186.
25. Mayer, R. E. Principles of multimedia learning based on social cues: Personalization, voice, and image principles. In *The Cambridge handbook of multimedia learning*, R. E. Mayer, Ed. Cambridge University Press, New York, NY, 2005, 201–212.
26. Mayer, R. E., and DaPra, C. S. An embodiment effect in computer-based learning with animated pedagogical agents. *Journal of Experimental Psychology: Applied 18*, 3 (2012), 239.
27. Mayer, R. E., and Moreno, R. A split-attention effect in multimedia learning: Evidence for dual processing systems in working memory. *Journal of educational psychology 90*, 2 (1998), 312.
28. Mayer, R. E., and Moreno, R. Nine ways to reduce cognitive load in multimedia learning. *Educational psychologist 38*, 1 (2003), 43–52.
29. Moore, M. G. Independent study. *Redefining the discipline of adult education 5* (1980), 16–31.
30. Moreno, R., Mayer, R. E., Spires, H. A., and Lester, J. C. The case for social agency in computer-based teaching: Do students learn more deeply when they interact with animated pedagogical agents? *Cognition and Instruction 19*, 2 (2001), 177–213.
31. Plass, J. L., Chun, D. M., Mayer, R. E., and Leutner, D. Supporting visual and verbal learning preferences in a second-language multimedia learning environment. *Journal of educational psychology 90*, 1 (1998), 25.
32. Reeves, B., and Nass, C. *How people treat computers, television, and new media like real people and places*. CSLI Publications and Cambridge university press, 1996.
33. Rosengrant, D., Hearrington, D., Alvarado, K., and Keeble, D. Following student gaze patterns in physical science lectures. In *AIP Conference Proceedings*, vol. 1413 (2012), 323.
34. Schmidt-Weigand, F., Kohnert, A., and Glowalla, U. A closer look at split visual attention in system-and self-paced instruction in multimedia learning. *Learning and Instruction 20*, 2 (2010), 100–110.
35. Soloman, B. A., and Felder, R. M. Index of learning styles. `http://www2.ncsu.edu/unity/lockers/users/f/felder/public/ILSdir/ILS-a.htm` (2004).
36. Sweller, J. Cognitive load theory, learning difficulty, and instructional design. *Learning and instruction 4*, 4 (1994), 295–312.
37. van Gog, T., and Scheiter, K. Eye tracking as a tool to study and enhance multimedia learning. *Learning and Instruction 20*, 2 (2010), 95–99.
38. Waldrop, M. M., and magazine, N. *Massive Open Online Courses, aka MOOCs, Transform Higher Education and Science*, March 2013.
39. Wang, N., Johnson, W. L., Mayer, R. E., Rizzo, P., Shaw, E., and Collins, H. The politeness effect: Pedagogical agents and learning outcomes. *International Journal of Human-Computer Studies 66*, 2 (2008), 98–112.