Instructional Video Design: Investigating the Impact of Monologue- and Dialogue-style Presentations

Bridjet Lee
Human Computer Interaction
Carleton University
Ottawa, Canada
bridjet.lee@carleton.ca

Kasia Muldner
Institute of Cognitive Science
Carleton University
Ottawa, Canada
kasia.muldner@carleton.ca

ABSTRACT
Instructional videos are frequently used in online courses and websites. Such videos may include an instructor delivering a monologue-style presentation, or alternatively, engaging in a dialogue with a student who appears in the video alongside of the instructor. We compared three instructional video designs (N = 77), including monologue and dialogue style presentations. To obtain a comprehensive view of the impact of video design, we used a variety of measures, including eye tracking data, learning gains, self-efficacy, cognitive load, social presence, and interest. Despite eye tracking data showing that participants in speaker-visible conditions spent significantly less time on the domain content, learning and related variables were similar in all three conditions, a result we confirmed with Bayesian statistics that provided substantial evidence for the null model. Altogether, we provide evidence that learning and interest are not enhanced by a dialogue-style presentation or visual presence of the instructor. However, further work is needed to investigate the effect of other domains, speaker persona and saliency, and configuration of the speakers in the instructional video.

Author Keywords
Instructional video design; monologue presentation; dialogue presentation; visual presence; eye tracking; interest

CCS Concepts
•**Applied computing** → **E-learning;** •**Human-centered computing** → *User studies;*

INTRODUCTION
Online learning continues to grow in popularity [3, 6, 15], thanks to the ubiquity of the Internet and of students who are eager to gain an education at their own pace and place. In addition to Massive Open Online Course (MOOC) platforms like Coursera and edX, most institutions now offer at least some of the content through online format. A popular way to deliver online education is through instructional videos. Given the popularity of this format and online learning in general, the question of how instructional videos should be designed is especially pertinent. Unfortunately, often the style choices for video content are based on "anecdotes, folk wisdom" [20], or precedent [21] instead of specific design decisions aimed at optimizing learning and/or affective outcomes like student interest.

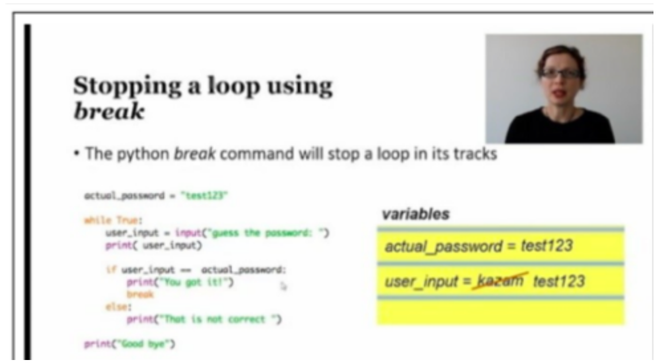

Figure 1. Instructional video using a monologue-style presentation, with the instructor shown in the video.

In the present study ($N = 77$), we compared learning and related outcomes from three instructional video formats. All three videos displayed the instructional content in a slide-based presentation, but the videos differed in terms of the style of narration related to that presentation and the visual presence of the instructor. Two of the videos used a *monologue*-style presentation, in which a solo instructor went over the instructional materials – in one video, the instructor could be seen (see Figure 1), while in the other, they could only be heard (i.e., they were not visually present in the video). In the third video, the instructor was shown having a *dialogue* with a student presented next to the instructor in the video (see Figure 2). While this latter format is less standard than a monologue-style presentation, there is some indication that students learn better from it [38]. To determine the effect of presentation style and visual presence, we collected a variety of measures, including eye tracking data to capture participants' visual attention, as well as learning and related measures like interest and cognitive load.

CHI '20, April 25–30, 2020, Honolulu, HI, USA.

ACM ISBN 978-1-4503-6708-0/20/04 ...$15.00.
http://dx.doi.org/10.1145/3313831.3376845

To date, work examining student learning from instructional videos has focused on the role of instructor presence, albeit with notable exceptions investigating the effect of presentation style (monologue vs. dialogue). Our work adds to this literature, as well as extends it. We tightly control for content by scripting the videos, making it possible to isolate effects if any to the video format. Some prior work has also scripted the video content [16, 13, 39], but has not collected affective and visual-attention measures via eye tracking. In general, our study is the first to provide evidence for equivalence of certain instructional video formats through Bayesian statistical methods. We describe our study and results after we present the related work.

RELATED WORK

Work most relevant to the present study can be broadly categorized as (1) studies that vary visual presence of the instructor and (2) studies that compare monologue- and dialogue-style presentation. Below, we describe each category; further details can be found in the *supplementary materials* that include a table of the key works and corresponding attributes, including the sample size, population, and key results and statistics reported.

Does it matter if the instructor is visible in the video?

Studies that evaluate the effect of instructor visual presence are motivated by social agency theory, which posits that social cues improve learning [33, 36]. Examples of beneficial social cues outside of the visual presence realm include polite instead of direct language [49], conversational wording instead of formal wording [34], and a human voice to narrate content instead of a mechanical (artificial) voice [36].

Another example of a social cue corresponds to the visual presence of the instructor in an instructional video (e.g., see Figure 1). Including the instructor can provide non-verbal social cues like facial expression and gesture, which has potential to motivate students as well as guide their attention. However, whether this occurs in practice is unclear, as the findings related to the effect of including the instructor in an instructional video are mixed. Some studies have found positive effects, in that students learned better from videos that showed the instructor as compared to videos where the instructor could be heard but not seen [10, 18, 48]. As an example, van Gog et al. [45] found that approximately one fifth of participants' fixations fell on the instructor's face when it was visible in an instructional video and that this helped learning: recall scores were higher than in the control condition where the instructor's face was not visible. The domain for this study was a physical puzzle being demonstrated by the instructor, and a proposed explanation for the benefit of physical presence was that participants who saw the instructor's face learned more because they were cued by the instructor's gaze to attend to the demonstration at relevant moments.

While arguments can be made for the benefits of showing the instructor, the visibility of the instructor also has the potential to distract observers, taking their attention away from the instructional content and/or splitting attention between that content and the instructor. In the broad context, principles of multimedia design propose minimizing distractors in the instructional materials as they increase cognitive load and impair learning [10, 23, 37, 35]. However, there is little evidence that the instructor's visual presence hinders learning: to our knowledge only one study has reported negative effects [50, Experiment 1], though this result was not replicated in subsequent experiments [50, Experiments 3 & 4].

The above-cited studies aside, the majority of work has reported null results regarding the impact on learning of an instructor's visual presence in an instructional video [7, 23, 29, 47, 46, 50]. To illustrate, in a recent study [46], participants were given either a video showing the instructor alongside instructional content (text and diagrams) or a video with only instructional content. Learning outcomes between the two groups were not significantly different (d = 0.35, 95% CI [-0.19; 0.89]). Several similar results influenced guidelines for multimedia design and in particular the image principle, which posits that the visual presence of the instructor does not necessarily contribute to better learning outcomes [33].

As the overview above highlights, there isn't overwhelming evidence that including the instructor in an instructional video helps learning, with various studies reporting null results, making it difficult to draw strong conclusions. Despite this, there may be other benefits of including the instructor, related to student interest and preferences. Indeed, several studies found that on average students preferred instructional videos with the instructor visible over videos where the instructor was not visible. In a longitudinal field study, Kizilcec et al. [28, Experiment 1] gave participants enrolled in a MOOC a choice to watch videos either with or without an inset video of the instructor's face (i.e., of the type shown in Figure 1, albeit in a different domain). Fifty seven percent of participants chose to watch videos with the instructor face, compared to 35% who chose to watch videos without the instructor. Further, participants who watched videos with the instructor's face perceived themselves as learning more, needing to exert less effort, and enjoying the experience more than did participants who watched videos without the face [28, Experiment 1]. In a second experiment, Kizilcec et al. [28] showed that the nature of instructor presence (constant vs. strategically shown) can affect cognitive load, social presence, and persistence. Wilson et al. [50] found similar positive perceptions of the instructor's visual presence in videos compared to videos that did not show the instructor — when asked to choose a format, a majority of participants cited the video with the instructor as the format they enjoyed and preferred and from which they learned best. As a third example, participants preferred having having the instructor visually present, despite the fact that this did not help them learn more than when the instructor was not visually present [29]. In contrast to these findings, however, other studies found that students view the instructor's visual presence as distracting [7], unhelpful [32], and cognitively demanding [23]. To illustrate, Homer at al. [23, Experiment 1] found that self-reported cognitive load was higher for participants who watched a video with the instructor visually present. Social presence — operationalized as the degree of social connection that participants felt towards the instructor — was also measured, and no conditional effects were found.

Do videos that include a dialogue between and instructor and their tutee promote better learning than a monologue?

The previous section reviewed work focusing on instructional videos in which a solo instructor presented a monologue-style exposition on the domain. An alternative strategy for instructional video design involves formatting the content as a dialogue instead of a monologue. With this design, the instructor discusses the content with a student (referred to as the *tutee* from now on, to distinguish from the student or participant observing the video). The studies that compare learning from a monologue vs. a dialogue do not manipulate visual presence, with the standard being that both speakers (instructor, tutee) are visible in the instructional video (e.g., see Figure 2 for an example).

One would expect that a naturally-occurring dialogue between an instructor and tutee would result in different content than a monologue delivered by the instructor and this is indeed the case. Earlier work focused on two features that had particular pedagogical potential: questions and misconceptions. Traditionally, in a dialogue, both questions and misconceptions would be produced by the tutee and addressed by the tutor. Compared to a monologue, in a dialogue there tend to be more questions and misconceptions because these are a byproduct of the presence of the tutee. Prior work manipulated the presence of these features by scripting the content of the instructional videos. Driscoll et al. [16] found that a dialogue-style video that contained deep-level reasoning questions resulted in better learning outcomes than a monologue-style video with no deep-level reasoning questions. However, once a monologue was scripted to include deep-level reasoning questions posed and answered by the instructor in the video, the learning outcomes between dialogue and monologue were similar [13, 19].

As far as misconceptions, Muller et al. [40] compared learning from an instructional video featuring a dialogue that included misconceptions, voiced by the tutee and refuted by the instructor in the video, against a monologue that did not include misconceptions. Participants who saw the dialogue video performed better on the posttest than participants who saw the monologue video. As was the case for questions, when a monologue video was scripted to include refuted misconceptions (posed as hypothetical errors by the instructor and subsequently refuted by the instructor), learning outcomes were comparable to ones obtained from the dialogue-style video, suggesting that it was the presence of the refuted misconceptions that enhanced learning, not the dialogue format itself [39].

Altogether, there is some evidence that the dialogue format alone is insufficient to produce more learning than a monologue format, since several studies have shown the benefits of dialogue disappear when content is scripted to equalize beneficial features (questions, misconceptions) between monologue- and dialogue-style videos. However, similar to the case made for the benefits of visual presence in a monologue-style video, dialogue-style videos may have positive benefits beyond learning outcomes.

One such benefit relates to self-efficacy. While studies comparing dialogue- vs. monologue-style videos have not reported on self-efficacy, seminal work on peer models using a monologue-style presentation showed that watching a peer in an instructional video make mistakes and demonstrate learning from the mistakes led children to feel more confident in their own skills than if they watched an instructor make mistakes [43]; confidence can be beneficial to learning because it encourages effort [5]. In [43], however, content was not equalized. Thus, it is an open question whether a dialogue video would improve self-efficacy if its content were equivalent to a monologue video.

A second potential benefit of dialogue-based videos, learning aside, is that they may promote student engagement in the instructional content. Several studies have found evidence of this increased engagement in the context of collaborative observation, where students watched an instructional video in pairs. In particular, a dialogue-based video promoted more discussion and/or substantive comments about the domain from the observers as compared to a monologue-style video [12, 38]. An explanation for this phenomenon is offered in [8], namely that the tutee in dialogue videos, as a peer model, attracts more observer attention than does the instructor. As a result, observers are influenced by and imitate the tutee's beneficial behaviours, such as making substantive comments and asking questions. Additionally, by watching the tutee struggle, make mistakes, and subsequently get corrected by the instructor, observers are hypothesized to empathize with the tutee and thus be motivated to learn [8].

PRESENT STUDY

As the review above highlights, studies investigating learning from instructional videos that show a monologue vs. dialogue have reported null effects on learning when the content of the videos was equalized between the two styles of presentation. As far as visual presence, studies that found benefits of including the instructor in the video have mainly used procedural and/or basic domains. None of these prior works have used Bayesian statistics (e.g., to provide evidence for the null hypothesis directly), nor have affective and visual attention measures been collected in a single study that varied instructor visual presence and style of presentation.

To fill these gaps, we compared learning from an instructional video that showed an instructor going over the material (i.e., a monologue, see Figure 1) to one that showed an instructor and a student going over the same instructional material (i.e., a dialogue, see Figure 2). As the control condition, we used a video that provided the same information but that did not show the instructor. We did not include a fourth condition where a dialogue could be heard but the speakers could not be seen as this has potential to be confusing due to the need to distinguish speakers, something that is facilitated by their visual presence.

The instructional videos presented a lesson on programming using the language Python. Programming is an activity that many students find difficult [11, 26], making it a suitable topic for the present study because it should provoke questions, misconceptions, and affective responses.

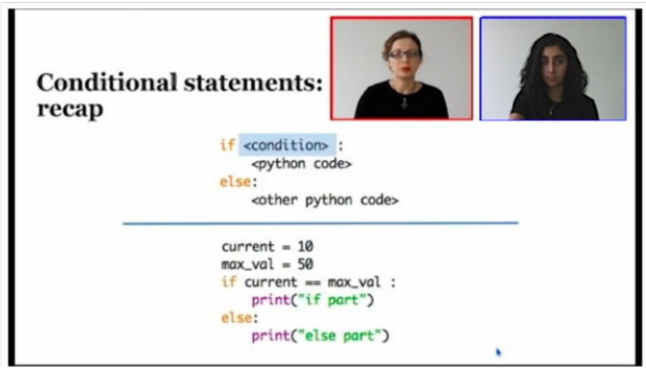

Figure 2. Screenshot of the dialogue-based video that showed the instructor and their tutee going over the content shown in the video together.

To obtain information on visual attention, operationalized by where the participants were looking in the video (e.g., instructor vs. the Python content shown in the video), we relied on eye tracking. Eye tracking data is beneficial because it can help explain why a given result occurred (e.g., if we found that participants in the monologue or dialogue conditions learned less than their peers in the control condition, and spent less time looking at the Python lesson in the video because they were looking at the instructor, this would provide evidence that the visual presence of the instructor was a distracting element). Learning was measured using pretest to posttest gains; related measures on factors like interest, cognitive load, and visual presence were also obtained using validated instruments. We had the following research questions:

- How does delivery format in an instructional video impact observers' visual attention?
- Does delivery format impact cognitive load and learning?
- What is the impact of delivery format on self-efficacy, interest toward the instructor, and interest toward programming?

Participants

Participants ($N = 77$, 52 female) were undergraduate students who were granted 2% bonus credit towards a course for participating in the study. The mean age was 19 and English was the primary language for 75% of participants. Most participants were Psychology majors. Participants were only eligible if they had not taken any university-level computer programming classes.

Apparatus & Materials

Apparatus. A desktop-mounted EyeLink 1000 eye tracker from SR Research was used to track eye movements. A chin rest was used to increase fidelity of data capture.

Instructional Videos. We created three instructional videos, one for each study condition: a monologue video that showed the instructor (see Figure 1), a dialogue video that showed the instructor and their tutee (see Figure 2), and a control video where the instructor could be heard but not seen (while technically this also presents a monologue, we refer to it as the control video to distinguish is from the monologue video that shows the instructor).

The three videos presented the same content based on a lesson about programming in Python. Because we recruited participants with limited or no programming knowledge, the coverage included fundamental programming concepts, e.g., variables, conditional statements, and basic while loops. To create the videos, we recorded a PowerPoint presentation and the instructor going over the content (as well as their tutee for the dialogue video), as we now describe.

The Python content in the three videos was displayed using a PowerPoint-based slideshow, and the same slideshow was used in the three videos. The slideshow presented Python code and brief facts about the code shown (e.g., that a break statement stops a loop, see Figure 2). The slideshow was composed of 17 slides with subtle animations used to highlight important elements.

The slides were narrated by the same instructor (control and monologue videos) and instructor and the tutee (dialogue video). The persona of the instructor was that of a personable expert who provided additional information about what was shown on a given slide. The tutee in the dialogue video was presented as a focused learner who asked questions, made substantive comments, and followed the lesson closely, but also demonstrated their role as a programming novice by occasionally verbalizing misconceptions, hesitations, and incorrect answers.

The content of what was said was scripted and the three videos all used the same script, so that what was said in the videos was equivalent. In particular, the same audio recording was used for the monologue and control video; for the dialogue video, the tutee was assigned the speaking role for some of the content. To control for speaking time in the dialogue video, the instructor and tutee were given roughly equal speaking time. The script for the narration included questions and refuted misconceptions, based on prior work showing these features promote learning [13, 39]. In the dialogue script, both the instructor and tutee asked questions to each other and provided explanations. In the monologue and control versions, the instructor asked and answered all questions as if addressing the viewer — this approach of instructor question-and-answer was used in prior work [13]. Misconceptions in this adjusted script were included as hypothetical assumptions that the instructor subsequently addressed: *"If you thought that, good guess, but it's actually because. . . "*. The three videos included the same number of questions and refuted misconceptions.

To create the three videos, we screen recorded the PowerPoint presentation displayed in *slideshow* mode, subsequently inserting the audio of the instructor (all conditions), and the videos of the speakers (the monologue and dialogue videos). To control for the position of the instructor in the dialogue video, half the participants in the dialogue condition saw a configuration with the instructor placed closer to the middle and the student to the right of them (see Figure 2), while the other half saw a reversed version with the student closer to the middle; no order effects were found. Cursor movements over the python content were included in all three videos to make it appear as if the speaker(s) were using a mouse to gesture to certain parts of the slideshow as they spoke. In the case of the dialogue video, external software was used to allow two mice to be connected and independently used on the same computer; this way, two cursors were visible on screen at the same time to show two users (the instructor and student) interacting with the slideshow.

To avoid fatiguing participants, we divided the presentation into two parts (part 1 = 10min; part 2 = 15-16 minutes depending on the condition, due to slight variation in speaking speed in the dialogue video as compared to the monologue and control videos). Each part included a review screen at the end summarizing the key points (the instructor/tutee were not shown on this review screen).

Questionnaires. To test Python knowledge, we developed a pretest and posttest (each was scored out of 19 points; the two tests were analogous, with superficial differences corresponding to variable names). We used established instruments to measure the following psychological variables:

- cognitive load using the nine-item instrument [31] measuring extraneous load (e.g., *'The instructions and explanations were full of unclear language'*), intrinsic load (e.g., *'The topics covered in the activity were very complex'*), and germane load (e.g., *'The activity really enhanced my understanding of the topics covered'*)
- interest in programming with the six-item questionnaire [42] (e.g., *'I think programming is interesting'*)
- social presence operationalized by interest in the instructor in the video [2]. In contrast to other social presence measures that focus on instructor/student interactions, this instrument was appropriate for our work because it did not assume interactivity in the learning environment, instead probing the learner's interest in the instructor with four items (e.g., *'The teaching style of the instructor held my interest'*)
- self-efficacy for programming with the seven-item instrument [22] (e.g., *'Even if I do not understand a programming problem at first, I am confident I will get it eventually'*)

For each of the variables, the target construct was measured by several items; to aggregate the items we followed the rubrics provided with the corresponding instruments.

Design & Procedure

We used a between-subjects design with three conditions based directly on the three instructional videos: control, monologue, dialogue. Participants in the *control* condition viewed the video where the instructor could be heard but not seen; the *monologue* group viewed the video that included the visual presence of the instructor (see Figure 1), and the *dialogue* group was given the video showing the instructor and the tutee (see Figure 2).

The procedure was the same for all three conditions. The study protocol was reviewed and approved by the university ethics board and participants consented to participation by signing a consent form. After signing the form, participants completed the Python pretest and self-efficacy questionnaire, and then were calibrated on the eye tracker. They were assigned to their condition in a round robin fashion. In each condition, participants first viewed part 1 of their instructional video (they could not pause or rewind, to control for time on task). They were then given a five minute break. After the break, participants were drift corrected on the eye tracker and recalibrated if needed. Participants watched part 2 of the video, and then filled in the post-intervention questionnaires (Python posttest, self-efficacy, interest, social presence, cognitive load).

RESULTS

The data were analyzed using two complimentary approaches: standard analysis via Null Hypothesis Significance Testing (NHST) and Bayesian analysis using the Bayes factor (*BF*). NHST and the associated *p* value is still the common standard for analysis, which is why we included results based on it. However, the Bayes factor approach has two distinct advantages: (1) it provides a measure of evidence present in the data for each model (alternative vs. null) and (2) it allows researchers to make claims about the null and alternative hypothesis [24]. In contrast, NHST only makes it possible to reject the null hypothesis in the case of a significant result and does not allow for claims regarding evidence for lack of an effect (i.e., a null model). Given these various considerations, there have been calls to present results from both frameworks (NHST, Bayesian), so that complimentary evidence can be compared [24, 41]. To obtain the Bayes factor, we used JASP [25].

Briefly, under the Bayesian approach, the "*likelihood of the data is considered under both the null and alternative hypotheses, and these probabilities are compared via the Bayes factor. The Bayes factor is a ratio that contrasts the likelihood of the data fitting under the null hypothesis with the likelihood of fitting under the alternative hypothesis.*"[24].[1] The ratio can be computed in either direction, and here, we report the Bayes factor as providing evidence for either the null model (no conditional differences) or the alternative model (conditional differences exist) as appropriate, stating which direction we are reporting. As the Bayes factor increases, it provides mounting evidence for the target model (either null or alternative, depending on the way the ratio is set up – Jasp allows the user to decide). Here, we follow the guidelines in [24] to interpret Bayes factors, as follows: *BF* = 1-3 provides anectodal evidence for the corresponding model; *BF* = 3-10 provides substantial evidence for the corresponding model; and *BF* > 100 provides decisive evidence (for full list of *BF* ranges and corresponding interpretations, see [24]).

We now present the results organized according to our three research questions.

[1] Bayesian analysis requires the specification of a prior but a so-called 'unfinromative prior' is advocated when there aren't clear prior beliefs. Since that is the case in our data, we used the Jasp defaults for the prior. This is the advocated approach since using specific priors can skew the results and requires clear justification [44].

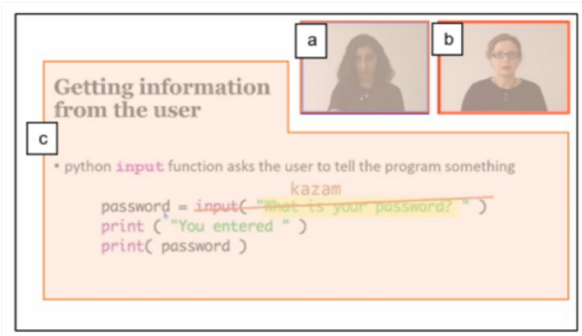

Figure 3. The AOIs used in the study. The dialogue condition included AOIs (a) and (b) while the monologue condition included only the instructor AOI (b); the python content AOI (c) was used in all three conditions.

How does delivery format in an instructional video impact observers' visual attention?

To obtain information on visual attention, we analyzed where participants were looking as they watched the instructional video. To do so, we relied on the eye tracking data (N = 68; data from the remaining nine participants was unusable due to technical failures; control n = 21, monologue n = 24, dialogue n = 23). We extracted data based on areas of interest (AOIs), as follows. A *python AOI* was drawn around the python content in the videos, see Figure 3c (this AOI was in the same location and of the same size in all three conditions, since all three relied on the same slideshow); the dialogue and monologue conditions additionally included AOIs around the speaker(s) (e.g., see (a) and (b) speaker AOIs present in the dialogue condition).

Visual attention was operationalized by dwell time and number of fixations in each AOI and condition, as reported by the eye tracker (because there was very little difference between parts 1 and 2 of each video on these variables, as we verified by visual inspection and inferential statistics, we collapsed across this variable by averaging the data across the two parts).

To account for slight variations between video length, as well as to provide an explicit measure of the proportion of content attended, we normalized the data (for dwell time, by dividing by total video time; for fixation count, by diving by the total number of fixations). Accordingly, we refer to the two dependent variables as *dwell time %* and *fixation %*.

Social agency theory predicts that the visual presence of the speaker in an instructional video should be beneficial. In order for this benefit to materialize, however, participants must look at the speaker(s). This was indeed the case in our study: Figure 4a shows that participants spend a moderate percentage of trial time, about 20%, looking at the instructor (monologue condition) and the instructor and tutee (dialogue condition).

Related to the social agency theory, there are proposals that observers of instructional videos will attend to the tutee more than the instructor [8], because they relate to the tutee more than the instructor. To check if this was the case in our study, we split the data for the dialogue condition according to visual attention to the instructor vs. the tutee. The results are in Figure 4b. In contrast to prior proposals, dialogue participants had a significantly higher fixation % on the instructor rather than tutee, $t(22) = 2.8$, $p = .01$, $d = 1.2$, and a marginally higher dwell time % on the instructor, $t(22) = 1.9$, $p = .07$. The Bayesian analysis confirmed there was substantial evidence in terms of higher fixation % on the instructor ($BF = 4.6$) but only weak evidence for the dwell time % variable being higher for the instructor ($BF = 1.1$).

Overall, the data shows that observers in the dialogue and monologue conditions visually attended to the speaker(s) in the instructional videos they watched. This did come at the cost of attention to the python content (see python AOI in Figure 2), descriptives are in Figure 4a. Specifically, participants spent less time looking at the python content in the dialogue and monologue conditions, as compared to the control. A between subjects ANOVA with condition as the independent variable confirmed this effect was reliable (dwell time %: $F(2, 65) = 11.76$, $p < .001$, $\eta^2 = .27$; fixation %: $F(2, 65) = 75.47$, $p < .001$, $\eta^2 = .70$).

Tukey's tests revealed that participants in the control condition spent a significantly higher percentage of trial time on the Python content than in the monologue condition (dwell time %: $p < .001$, $d = 1.01$; fixation %, $p < .001$, $d = 2.94$) and dialogue condition (dwell time %: $p < .001$, $d = 1.26$; fixation %: $p < .001$, $d = 1.26$). Compared to the dialogue condition, the monologue condition had more fixations on the python content ($p = .05$) but dwell time was not significantly different.

Analysis with Bayes factor confirmed the above results related to attention to python content. The estimated Bayes factor indicated the data were 499.2 times more likely to occur under a model including effect of video type than a model without it (i.e., the null model). Posthoc analyses mirrored the above results in that there was strong evidence that the control condition devoted the most attention to the python content ($BF = 24.1$ and $BF = 176.2$ for control-monologue and control-dialog comparisons, in favor of the alternative model), but there was weak evidence regarding the dialogue-monologue comparison ($BF = .5$), indicating this may not be a reliable effect.

Participants in the monologue and dialogue conditions did not make up reduced time on the python content by spending longer on the review screen presented at the end of part 1 and part 2 videos in all three conditions. In particular, the effect of condition on review time was not significant, $F(2, 65) = 1.05$, $p = .36$, $\eta^2 = .03$ (review time is raw time spent viewing the review slide). The obtained Bayes factor indicated that there was substantial evidence for the null model ($BF = 3.7$), confirming the NHST results.

Summary. Participants did look at the speakers in the dialogue and monologue conditions, which could have beneficial effects like improved learning and/or interest based on social cues provided by the speakers. On the other hand, attention to the speakers reduced attention to the python content; time that was not made up on the review screen. Moreover, participants in the monologue and dialogue conditions had to divide their attention between the speaker(s) and the learning content.

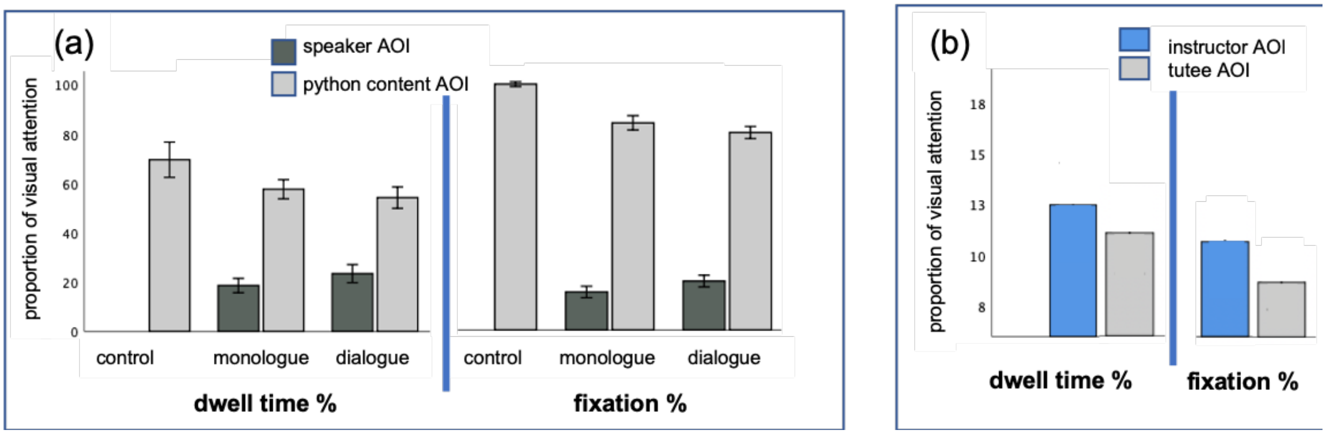

Figure 4. (a) Mean dwell time % and fixation % showing visual attention to the speakers (instructor in the monologue condition; instructor and tutee in the dialogue condition; N/A for the control condition as the instructor was not visible) vs. the python content available available in all three conditions. Error bars represent 95% confidence intervals. (b) Mean dwell time % and fixation % on the instructor AOI vs. the tutee AOI in the dialogue condition, showing attention to the instructor vs. the tutee (error bars not shown as the comparison was on the difference scores).

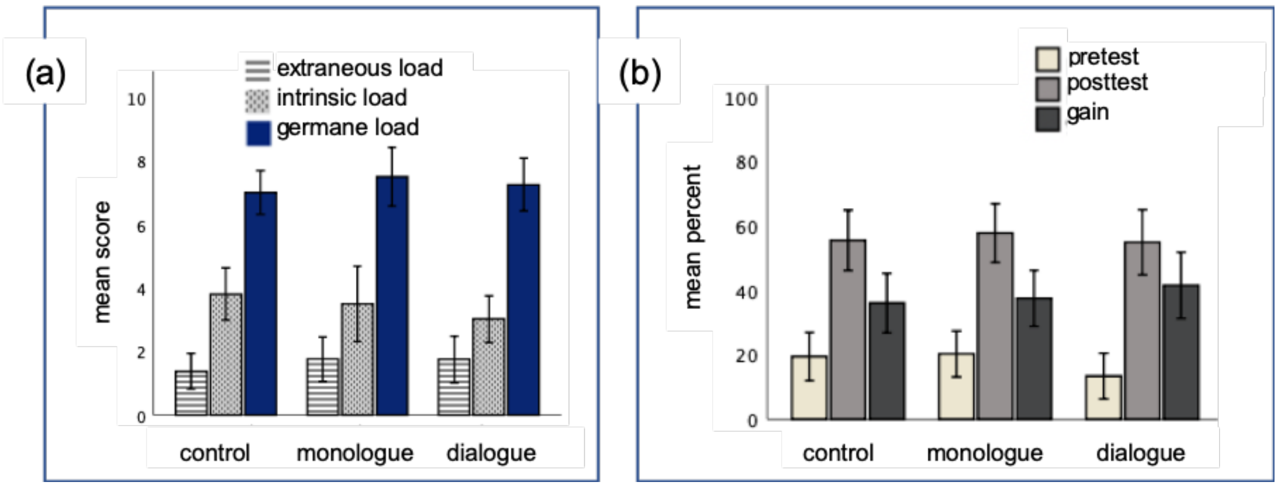

Figure 5. (a) Self-reported cognitive load in each condition (max is 10); (b) Mean pretest %, posttest %, and gain (posttest - pretest) % in each condition. Error bars represent 95% confidence intervals.

How did these factors affect cognitive load and learning in the three conditions? We address this question next.

Does delivery format impact cognitive load and learning?
The following analysis is based on data from all 77 participants (control $n = 26$; monologue $n = 26$; dialogue $n = 25$). We begin with the impact of instructional video design on cognitive load. We report on the three types of load (extraneous, intrinsic, germane) shown in Figure 5a, but of primary interest is extraneous load, which is cognitive load imposed by the design of the learning material, as we were interested in whether the presence of the speakers, and attention to them, increased this type of cognitive load. There were no significant differences between conditions in self-reported extraneous load, $F(2, 74) = .46$, $p = .63$, $\eta^2 = .01$ and the Bayes Factor confirmed there was substantial evidence for the null model ($BF = 6.2$). The results for the two other types of load mirror this pattern (intrinsic load: $F(2, 74) = .75$, $p = .48$, $\eta^2 = .02$, Bayes Factor indicated substantial evidence for the null model, $BF = 5.03$; germane load: $F(2, 74) = .39$, $p = .67$, $\eta^2 = .01$, Bayes Factor indicated substantial evidence for the null model, $BF = 6.6$).

Thus, we have evidence that the presence of the speakers did not increase cognitive load. Given this, we would expect that learning would not be diminished in the monologue and dialogue conditions, but would it be increased due to the presence of the instructor and/or tutee? The learning results addressing this question are below.

Learning was measured using the standard method of calculating gains from pretest to posttest (*gain* = *posttest* – *pretest*; the maximum achievable score on each test was 19 points - for interpretability, we show scores as percentages. Participant pretest and posttest scores are shown in Figure 5b. Participants began the experiment with similar levels of programming knowledge prior to the intervention as indicated by an ANOVA with pretest score as the dependent variable, $F(2, 74) = 1.13$, $p = .33$. As expected, the pretest scores were low

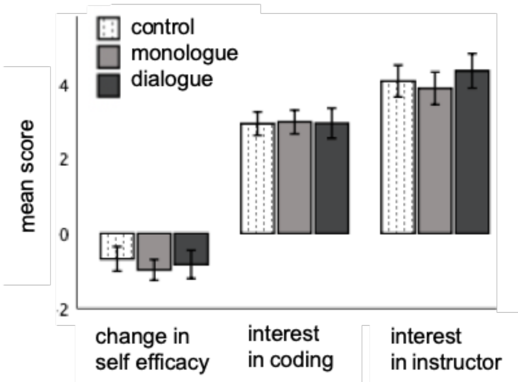

Figure 6. Mean self-reported (i) change in self efficacy (post-pre), (ii) interest in programming post intervention, (iii) interest the instructor shown in the video. Error bars represent 95% confidence intervals.

(around 20%), as we recruited participants with little or no programming background.

Collapsing across conditions, participants did improve significantly from pretest to posttest, $t(76) = 14.7$, $p < .001$, $d = 1.7$, with substantial evidence for the alternative model (BF = 6.3e20), indicating that learning occurred. However, the type of video watched did not differentially impact gain scores, which were similar across the three conditions (see Figure 5). A between subjects ANOVA with condition as the independent variable did not find a significant effect of condition on gain scores, $F(2, 74) = .38$, $p = .69$, $\eta^2 = .01$. The estimated Bayes factor indicated the data were 6.7 times more likely to occur under the null model, providing substantial evidence that video type did not have an effect on gain scores.

What is the impact of delivery format on self-efficacy and interest toward programming?

Overall, participants paid attention to the speakers in the dialogue and monologue videos but this did not impact learning. Despite this lack of learning, an argument could still be made in favor of designing instructional videos that included a dialogue between an instructor and their tutee if this design inspired higher self-efficacy towards programming and/or interest than the other conditions. As the descriptives in Figure 6 show, this did not turn out to be the case, with little differences between the three experimental conditions.

There was no conditional effects on changes in self-reported self-efficacy towards programming from pre to post intervention, $F(2, 74) = .85$, $p = .43$, $\eta^2 = .02$. The estimated Bayes factor indicated the data were 4.6 times more likely to occur under the null model, providing substantial evidence that video type did not have an effect on self-efficacy. The self-efficacy decreased from pretest to posttest, possibly because participants were better able to estimate their knowledge of programming after being introduced to it (i.e., self calibration is improved with expertise).

Along a similar vein, there were no significant differences between conditions in terms of how interested participants felt towards programming after the intervention, $F(2, 74) = .02$, $p = .98$, $\eta^2 < .001$; this result was replicated with the Bayesian analysis ($BF = 8.8$, indicating the null model was almost 9 times more likely than the alternative model). Format of the instructional video also did not affect how interested participants felt towards the instructor in the video, $F(2, 74) = 1.20$, $p = .31$, $\eta^2 = .03$, despite the fact that they were only visually present in the dialogue and monologue conditions; this result was again confirmed with the Bayesian analysis ($BF = 3.6$).

DISCUSSION AND FUTURE WORK

Our work addressed two high level questions: (1) Does a dialogue style presentation in an instructional video improve outcomes over a monologue-style presentation? (2) In the context of a monologue-style video, is the visual presence of the instructor beneficial? The answer to both these questions based on our results is "no". This was not because students didn't pay attention to the speakers in the instructional videos: Our eye tracking analysis provided evidence that participants looked at the speakers in the two instructional videos that showed them (monologue, dialogue conditions) for about one fifth of the total trial time. Despite this clear pattern of attention to the speakers, however, we found no differences in learning. This finding is notably similar to that of Kizilcec et al. [29], who found that participants spent 41% of trial time looking at the instructor when they were present, without it significantly influencing their learning. We also did not find instructor presence had an effect on feelings of social presence, interest, and cognitive load. The same pattern emerged when we compared the dialogue to the monologue condition, suggesting that the presence of the tutee in the video did not have an effect on learning or related outcomes. These results were confirmed with Bayesian analysis that provided evidence for the null model.

Effect of a dialogue-style presentation: Synthesizing the findings

Prior work comparing the effect of dialogue vs. monologue style presentation used learning as the primary measure, without including eye tracking data or affective outcomes as we did in the present study. What does this work say about the effect of presentation style on learning? The answer depends on whether the narration around the content in the instructional videos was scripted to equalize it between the dialogue and monologue conditions.

Prior studies showed that when the narration in the instructional video was not scripted, meaning that there were differences in what was said in the dialogue and monologue videos, students learned better from dialogue than monologue-based videos [40, 13, 38]. The caveat is that this benefit disappeared when both types of videos were scripted to include the same features conjectured to improve learning (refuted misconceptions, questions) [40, 13]. Thus, in general our findings agree with the former works showing that the benefits of dialogue videos disappear when the content is scripted, with our work adding information on interest, cognitive load, and self-efficacy.

What hasn't been clear to date is whether, after controlling for content, students find dialogue videos more interesting or less cognitively demanding, which could be an argument for designing them. In our study we did control for video content, and we did not find evidence that the dialogue-based video improved learning, interest, or related variables like self-efficacy and cognitive load over a monologue presentation. We now discuss why dialogue could inspire interest and speculate on reasons why we did not see evidence of this in our study.

One reason that a dialogue video could inspire more interest than a monologue video is due to the presence of the tutee, which provides a peer model that observers can relate to. In fact, prior analysis of verbal protocols found that observers made more references to the tutee in the video than the instructor [8]. In contrast, in our study, participants paid *less* visual attention to the tutee in the video than the instructor. There are a number of possible explanations for this discrepancy in findings. One is that the tutee in our study spoke less than the tutee in the video in [8] – since speaking attracts attention, this could explain why less attention would be payed to the tutee in our study. This explanation is unlikely because we scripted the content to roughly equalize speaking time between the instructor and the tutee, while in [8], content was unscripted. When tutorial dialogue is not scripted, instructors tend to speak more than the their tutees [9]. A second potential explanation is that attention is operationalized differently - in our study, by dwell and fixation time based on eye tracking data, while in [8], by verbal references to the speakers. Thus, it may be that observers spent less time looking at the tutee as compared to the tutor in both studies - something that can't be determined as an eye tracker was not used in [8]. A third explanation relates to what the tutee was doing in our study vs. in [8]: our tutee was discussing the content with the instructor, but not actively solving problems, while in [8], the tutee was at a whiteboard working on a problem with the tutor. Perhaps this more active style of presentation affected attention devoted by the observers to the tutee.

A related factor that could potentially boost the benefit of dialogue is the activity level of the observers. In the vast majority of studies comparing learning from a dialogue vs. a monologue presentation in instructional videos, the observers passively watched the videos without doing anything [40, 39, 16, 13]. It may be that the benefits of dialogue do not emerge without an interactive component requiring the observers to apply what they learned from the video. In the Muldner et al. study [38] that did find a benefit of a dialogue presentation, observers completed a worksheet in pairs while watching an instructional video featuring the same worksheet being completed by the instructor (monologue condition) or instructor and their tutee (dialogue condition). The observers in the dialogue condition learned more than ones in the monologue condition and one explanation for why they did is because they were more constructive. Evidence for this conjecture came from the qualitative analysis showing that dialogue-video observers made more substantive contributions and so were less passive than the monologue-video observers. However, this work did not script the content of the videos and the dialogue videos had more beneficial features (refuted misconceptions, questions) than the monologue videos. Since these features improve learning, it is an open question as to whether a dialogue format would improve observer learning from constructive activities while watching the video, over a monologue format, once these the content of the videos is equalized. This question awaits future investigation.

Yet another factor that may have affected our results pertains to how the instructor and tutee were positioned in the instructional videos. In the present study, they were shown in separate frames. We chose this setup so the speakers would be facing the viewer with the aim of engaging the viewer, but this type of setup may have had the disadvantage of making the speakers' interaction appear less natural than if they were facing each other in a single frame.

Visual presence of the instructor

The visual presence of the instructor in the monologue condition, and the instructor and tuttee in the dialogue condition, had potential to add visual social cues to the videos. These social cues — namely facial expression, gestures (nodding, head tilting), and gaze — were predicted, on one hand, to arouse participants' feelings of social presence and interest in the content, thereby facilitating learning. On the other hand, a competing prediction was that the visual presence of the speaker(s) would distract away from the python content, induce greater cognitive load, and so impair learning.

It is somewhat striking that monologue and dialogue participants spent significantly less attention on the learning content than control participants but did not learn less than the control group that did not see the instructor in the videos. However, there is some precedent for this in prior work investigating the effect of visual presence in a monologue-style presentation. Specifically, van Wermeskerken et al. [46] found similar results when comparing instructional videos: while participants in an instructor-absent condition spent about 79% of trial time on the learning content compared to 52% for participants in an instructor-present condition, no significant learning differences emerged. Similarly, van Wermeskerken and van Gog [47] found high proportions of trial time spent on learning content — roughly 94% in an instructor-absent condition and 81% in two instructor-present conditions — without any corresponding differences in learning outcomes.

In the broader context including studies without eye tracking, the learning results mirror ours in that null effects from NHST are often reported [7, 23, 32, 50]. However, learning may be affected if participants were to pay more attention to the instructor (e.g., increased if the students are motivated by the instructor and attention is not needed on the learning content, or diminished if attention is needed on the learning content but paid to the instructor instead). A factor that likely plays a role in how much attention is devoted to the speaker(s) is instructor persona (and tutee persona if they are present), including how gregarious and animated are the speaker personas. Instructor enthusiasm is hypothesized to increase student attention and [4] and has been shown to be positively correlated with student learning [27]. In general, salient features like bright colors [14], in the present context expressed for instance through clothing or jewelry, as well as sudden movements [1], are also important predictors of attention. Thus, investigating instructor salience is another future avenue of research.

The above-cited studies did not include the additional measures used in our study related to interest, cognitive load, etc. Colliot and Jamet (2018) did measure social presence, interest, and cognitive load as we did, and found the same null effects of speaker visual presence in all measures (with traditional NHST testing). One limitation of these measures is that they are self reports, which are not always accurate. Thus, there is interest in developing more objective ways to measure constructs like cognitive load, for instance using eye tracking technologies [30, 17].

Conclusion

The rise in popularity of online learning platforms that rely on instructional videos makes it particularly important to investigate how to design instructional videos that inspire learning and interest. Here, we manipulated the style of presentation (and the visual presence of the instructor in the control video). While the visual presence of the speaker(s) drew participants' attention at the expense of attention to the learning content, learning and other psychological measures were not impacted and were comparable across three instructional video designs (monologue-style video, dialogue-style video, and a control video where the instructor's monologue could be heard but the instructor was not seen). Bayesian statistical methods allowed us to provide evidence for the null model, extending prior work relying on NHST.

While our study adds to the mounting evidence that learning from a dialogue-style presentation is not better than a monologue one once content is equalized, and that speaker visual presence doesn't have an effect on key learning variables, it would be premature in our opinion to conclude that that these factors do not have an impact in all contexts. For instance, a context in which the observers are actively applying the material presented in the video, such as completing a worksheet, may change this pattern of results, and thus should be addressed by future work. Moreover, the generalizability of our findings is limited to the present domain, namely programming. While this domain requires visual processing, the content of a given slide was fairly static. It may be the case that domains that require more visual processing would be differently affected by the visual presence of the instructor (and tutee) than domains that require less visual processing. This, along with other limitations outlined above, points to the need for additional research.

ACKNOWLEDGEMENTS

We thank Sara Nafar for her role as the tutee in the instructional videos, as well as the anonymous reviewers for their constructive comments. This work was supported with an NSERC Discovery Grant #1507.

.

REFERENCES

[1] Richard A. Abrams and Shawn E. Christ. 2003. Motion Onset Captures Attention. *Psychological Science* 14, 5 (2003), 427–432.

[2] A. B. Adcock and R. N. Van Eck. 2005. Reliability and factor structure of the Attitude Toward Tutoring Agent Scale (ATTAS). *Journal of Interactive Learning Research* 16, 2 (2005), 195–212.

[3] I. E. Allen and J. Seaman. 2017. *Digital learning compass: Distance education enrollment report 2017*. Technical Report. Babson College/eLiterate/WCET, Wellesley.

[4] Elisha Babad. 2007. *Teachers' Nonverbal Behavior and its Effects on Students*. 201–261.

[5] Albert Bandura. 1982. Self-efficacy mechanism in human agency. *American Psychologist* 37, 2 (02 1982), 122–147.

[6] Tony Bates. 2018. The 2017 national survey of online learning in Canadian post-secondary education: methodology and results. *International Journal of Educational Technology in Higher Education* 15, 29 (2018).

[7] Eta S Berner and Barbara Adams. 2004. Added value of video compared to audio lectures for distance learning. *International Journal of Medical Informatics* 73, 2 (2004), 189 – 193.

[8] Michelene T. H. Chi, Seokmin Kang, and David L. Yaghmourian. 2017. Why students learn more from dialogue- than monologue-videos: Analyses of peer interactions. *Journal of the Learning Sciences* 26, 1 (2017), 10–50.

[9] Michelene T. H. Chi and Marguerite Roy. 2010. How Adaptive Is an Expert Human Tutor?. In *Intelligent Tutoring Systems*, Vincent Aleven, Judy Kay, and Jack Mostow (Eds.). Springer Berlin Heidelberg, Berlin, Heidelberg, 401–412.

[10] Tiphaine Colliot and Éric Jamet. 2018. Understanding the effects of a teacher video on learning from a multimedia document: An eye-tracking study. *Educational Technology Research and Development* 66, 6 (01 Dec 2018), 1415–1433.

[11] C. Connolly, E. Murphy, and S. Moore. 2009. Programming anxiety amongst computing students – A key in the retention debate? *IEEE Transactions on Education* 51, 1 (2009), 52–56.

[12] Scotty D. Craig, Michelene T. H. Chi, and Kurt VanLehn. 2009. Improving classroom learning by collaboratively observing human tutoring videos while problem solving. *Journal of educational psychology* 101, 4 (11 2009), 779–789.

[13] Scotty D. Craig, Jeremiah Sullins, Amy Witherspoon, and Barry Gholson. 2006. The Deep-Level-Reasoning Question Effect: The Role of Dialogue and Deep-Level-Reasoning Questions During Vicarious Learning. *Cognition and Instruction* 24, 4 (2006), 565–591.

[14] Stefan Van der Stigchel, Artem V. Belopolsky, Judith C. Peters, Jasper G. Wijnen, Martijn Meeter, and Jan Theeuwes. 2009. The limits of top-down control of visual attention. *Acta Psychologica* 132, 3 (2009), 201 – 212.

[15] Tricia Donovan, Tony Bates, Jeff Seaman, Denis Mayer, Éric Martel, Ross Paul, Brian Desbiens, Vivian Forssman, and Russ Poulin. 2019. *Tracking Online and Distance Education in Canadian Universities and Colleges: 2018*. Technical Report. Canadian Digital Learning Research Association.

[16] David M. Driscoll, Scotty D. Craig, Barry Gholson, Matthew Ventura, Xiangen Hu, and Arthur C. Graesser. 2003. Vicarious Learning: Effects of Overhearing Dialog and Monologue-like Discourse in a Virtual Tutoring Session. *Journal of Educational Computing Research* 29, 4 (2003), 431–450.

[17] Andrew T. Duchowski, Krzysztof Krejtz, Izabela Krejtz, Cezary Biele, Anna Niedzielska, Peter Kiefer, Martin Raubal, and Ioannis Giannopoulos. 2018. The Index of Pupillary Activity: Measuring Cognitive Load Vis-à-Vis Task Difficulty with Pupil Oscillation.

[18] Q. Dunsworth and R. K. Atkinson. 2007. Fostering multimedia learning of science: Exploring the role of an animated agent's image. *Computers & Education* 49, 3 (2007), 677–690.

[19] B. Gholson, A. Witherspoon, B. Morgan, J. K. Brittingham, R. Coles, A. C. Graesser, J. Sullins, and S. D Craig. 2009. Exploring the deep-level reasoning questions effect during vicarious learning among eighth to eleventh graders in the domains of computer literacy and Newtonian physics. *Instructional Science* 37, 5 (2009), 487–493.

[20] Philip J. Guo, Juho Kim, and Rob Rubin. 2014. How Video Production Affects Student Engagement: An Empirical Study of MOOC Videos. In *Proceedings of the First ACM Conference on Learning @ Scale Conference (L@S '14)*. ACM, New York, NY, USA, 41–50.

[21] A. Hansch, C. Newman, L. Hillers, T. Schildhauer, K. McConachie, and P. Schmidt. 2015. Video and online learning: Critical reflections from the field. *HIIG Discussion Paper Series* 2015, 2 (2015), 1–34.

[22] Helen M. Hendy, Nancy Schorschinsky, and Barbara Wade. 2014. Measurement of math beliefs and their associations with math behaviors in college students. *Psychological assessment* 26, 4 (12 2014), 1225–1234.

[23] B. Homer, J. Plass, and L. Blake. 2008. The effects of video on cognitive load and social presence in multimedia-learning. *Journal of Educational Psychology* 24, 3 (2008), 786–797.

[24] Andrew F. Jarosz and Jennifer Wiley. 2014. What Are the Odds? A Practical Guide to Computing and Reporting Bayes Factors. *Journal of Problem Solving* 7, 1 (2014).

[25] JASP Team. 2019. JASP (Version 0.10.2)[Computer software]. (2019). `https://jasp-stats.org/`

[26] Tony Jenkins. 2002. On the difficulty of learning to program. In *Proc. 3rd Ann. Conf. HEA Learning Teaching Support Netw.* 53–58.

[27] Melanie Keller, Anita Hoy, Goetz Thomas, and Anne Frenzel. 2015. Teacher Enthusiasm: Reviewing and Redefining a Complex Construct. *Educational Psychology Review* 28 (12 2015).

[28] René F. Kizilcec, Jeremy N. Bailenson, and Charles J. Gomez. 2015. The instructor's face in video instruction: Evidence from two large-scale field studies. *Journal of Educational Psychology* 107, 3 (08 2015), 724–739.

[29] René F. Kizilcec, Kathryn Papadopoulos, and Lalida Sritanyaratana. 2014. Showing face in video instruction: Effects on information retention, visual attention, and affect. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '14)*. ACM, New York, NY, USA, 2095–2102.

[30] Thomas Kosch, Mariam Hassib, Paweł W. Woundefinedniak, Daniel Buschek, and Florian Alt. 2018. Your Eyes Tell: Leveraging Smooth Pursuit for Assessing Cognitive Workload. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems (CHI '18)*. Association for Computing Machinery, New York, NY, USA, Article Paper 436, 13 pages.

[31] Jimmie Leppink, Fred Paas, Cees P. M. Van der Vleuten, Tamara Van Gog, and Jeroen J. G. Van Merriënboer. 2013. Development of an instrument for measuring different types of cognitive load. *Behavior Research Methods* 45, 4 (01 Dec 2013), 1058–1072.

[32] A. Lyons, S. Reysen, and L. Pierce. 2012. Video lecture format, student technological efficacy, and social presence in online courses. *Computers in Human Behaviour* 28 (2012), 181–186.

[33] R. E. Mayer. 2014. *Principles based on social cues in multimedia learning: Personalization, voice, image, and embodiment principles* (2 ed.). Cambridge University Press, New York, NY, 345–368", booktitle = "The Cambridge Handbook of Multimedia Learning.

[34] Richard E. Mayer, Sherry Fennell, Lindsay Farmer, and Julie Campbell. 2004. A personalization effect in multimedia learning: Students learn better when words are in conversational style rather than formal style. *Journal of Educational Psychology* 96, 2 (06 2004), 389–395.

[35] Richard E. Mayer and Roxana Moreno. 1998. A Split-Attention Effect in Multimedia Learning: Evidence for Dual Processing Systems in Working Memory. *Journal of Educational Psychology* 90, 2 (1998), 312–320.

[36] R. E. Mayer, K. Sobko, and P. D. Mautone. 2003. Social cues in multimedia learning: Role of speaker's voice. *Journal of Educational Psychology* 95, 2 (06 2003), 419–425.

[37] R. Moreno, R. E. Mayer, H. A. Spires, and J.C. Lester. 2001. The case for social agency in computer-based teaching: Do students learn more deeply when they interact with animated pedagogical agents? *Cognition and Instruction* 19, 2 (2001), 177–213.

[38] Kasia Muldner, Rachel Lam, and Michelene T. H. Chi. 2014. Comparing Learning from Observing and from Human Tutoring. *Journal of Educational Psychology* 106, 1 (2014), 69–85.

[39] D. A. Muller, J. Bewes, M. D. Sharma, and P Reimann. 2008. Saying the wrong thing: improving learning with multimedia by including misconceptions. *Journal of Computer Assisted Learning* 24 (2008), 144–155.

[40] Derek A. Muller, Manjula D. Sharma, John Eklund, and Peter Reimann. 2007. Conceptual change through vicarious learning in an authentic physics setting. *Instructional Science* 35, 6 (01 Nov 2007), 519–533.

[41] Daniel S. Quintana and Donald R. Williams. 2018. Bayesian alternatives for common null-hypothesis significance tests in psychiatry: a non-technical guide using JASP. *BMC Psychiatry* 18, 1 (07 Jun 2018), 178.

[42] J. Rotgans and H. Schmidt. 2011. Situational interest and academic achievement in the active-learning classroom. *Learning and Instruction* 21, 1 (2011), 58–67.

[43] Dale H. Schunk and Antoinette R. Hanson. 1985. Peer models: Influence on children's self-efficacy and achievement. *Journal of educational psychology* 77, 3 (06 1985), 313–322.

[44] Johnny van Doorn, Don van den Bergh, Udo Bohm, Fabian Dablander, Koen Derks, Tim Draws, Alexander Etz, Nathan J Evans, Quentin F Gronau, Max Hinne, and et al. 2019. The JASP Guidelines for Conducting and Reporting a Bayesian Analysis. (Jan 2019).

[45] T. van Gog, I. Verveer, and L. Verveer. 2014. Learning from video modeling examples: Effects of seeing the human model's face. *Computers & Education* 72 (2014), 323–327.

[46] M. van Wermeskerken, S. Ravensbergen, and T. van Gog. 2018. Effects of instructor presence in video modeling examples on attention and learning. *Computers in Human Behavior* 89 (2018), 430–438.

[47] M. van Wermeskerken and T. van Gog. 2017. Seeing the instructor's face and gaze in demonstration video examples affects attention allocation but not learning. *Computers & Education* 113 (2017), 98–112.

[48] J. Wang and P. D. Antonenko. 2017. Instructor presence in instructional video: Effects on visual attention, recall, and perceived learning. *Computers in Human Behavior* 71 (2017), 78–89.

[49] Ning Wang, W. Lewis Johnson, Richard E. Mayer, Paola Rizzo, Erin Shaw, and Heather Collins. 2008. The politeness effect: Pedagogical agents and learning outcomes. *International Journal of Human-Computer Studies* 66, 2 (2008), 98 – 112.

[50] K. Wilson, M. Martinez, C. Mills, S. D'Mello, D. Smilek, and E. Risko. 2018. Instructor presence effect: Liking does not always lead to learning. *Computers & Education* 122 (2018), 205–220.