An Eye Tracking Comparison of Instructional Videos Showing a Monologue Versus a Dialogue: Impacts on Visual Attention, Learning, and Psychological Variables

by

Bridjet Jennifer Lee

A thesis submitted to the

Faculty of Graduate and Postdoctoral Affairs

in partial fulfillment of the requirements for the degree of

Master of Arts in Human-Computer Interaction

Carleton University

Ottawa, Ontario

August, 2019

Abstract

The present study aimed to synthesize two disparate domains of instructional video research to investigate what impacts occurred from, on one hand, the visual presence of the speaker(s), and on the other hand, the format of a dialogue. Seventy-seven participants watched either a narrated control video without the instructor visible, a monologue video with the instructor visible, or a dialogue video between an instructor and student, both visible. To compare the conditions, we examined learning outcomes, visual attention, self-efficacy, mindset, cognitive load, social presence, and interest. Despite eye tracking data showing that participants in speaker-visible conditions spent significantly less time attending to the learning content, we found no conditional differences on measures of learning, social presence, cognitive load, self-efficacy, or mindset. These results suggest that neither speaker visual presence nor dialogue format affected learning or participants' perceptions of the videos.

Acknowledgments

I would like to thank the countless people who supported, guided, and cheered me along my journey through this work. This project would have certainly never left the ground if not for Prof. Kasia Muldner, whose supervision cleared my path of many obstacles and brought invaluable direction and clarity to my ideas and writing. Thank you to Sara for volunteering to be on camera as our intrepid 'student', and to the participants for contributing their time and feedback. Thank you to the faculty, fellow students, staff, and others that I had the pleasure of meeting through the Carleton HCI program — being surrounded by the wealth of knowledge and passion for this field was a constant reminder of the meaningful challenge of awesome, empathetic, human-centred design.

To my UOSalsa familia, thank you for bringing so much joy into my life and helping me nurture my love of dance. A regular dose of salsa, bachata, and kizomba was the perfect foil to long days spent reading and writing.

Finally, thank you to my friends and family for encouraging me to pursue my dreams and for being there through both stress and celebration. And to Curtis, for being ever steadfast in love, support, and LaTeX edits.

Table of Contents

List of Tables

List of Figures

Chapter 1

Introduction

Online learning continues to push the boundaries of the traditional classroom, thanks to the ubiquity of the Internet and of learners who are eager to gain an education at their own pace and place. Massive Open Online Course (MOOC) platforms like Coursera and edX, as well as online courses in higher education institutions, make attractive options for large scores of learners to access learning content. Such demand for easy access to high-quality learning imposes pressure on content providers to create and distribute material that not only rivals in-person instruction, but is also cost-effective to produce.

Instructional videos are commonly the solution to this challenge in online learning (Hansch et al., 2015). However, what remains unclear is how to design video content to meet the learning needs of students on a broad scale. Often, the style choices for video content are based simply on "anecdotes, folk wisdom" (Guo et al., 2014, p. 42), or precedent (Hansch et al., 2015). Though video production styles present myriad ways to effectively convey information and motivate learners, they must be applied judiciously by designers or risk losing their efficacy (Koumi, 2006). Given this problem, there is interest in identifying factors that promote learning from instructional videos as well as related factors like student self-efficacy and interest.

A common approach for designing instructional videos is the narrated slideshow, where an instructor provides monologue-style narration over slides but does not appear in the slideshow. A variation of this approach includes a video of the instructor on the slides being narrated. Each of these approaches are similar in that the instructor delivers a *monologue* about the instructional materials, but have been argued to differ in the specific benefits and detriments they offer, which will be elaborated upon below. Studies comparing conditions where the instructor is presented versus not are referred to here as *visual presence studies*. Another less-explored approach to instructional video design involves having both an instructor and a student engage in a *dialogue* to present learning content, typically also displayed alongside a slideshow. Again, prior *dialogue studies* have examined the ways in which such a presentation style using two speakers affects learners both positively and negatively. In response to a lack of clear evidence and the ongoing ubiquity of online learning, current work continues to investigate the effects of various instructional video formats to better understand their impacts on learning and related variables like attention, beliefs, and affect.

The present thesis aimed to contribute to this understanding by comparing the three video formats to each other on a wide variety of measures. Learning was the main outcome of interest, but other variables such as visual attention and self-efficacy were selected because of their potential to predict or explain learning and affect. Ultimately, such data can be used to inform the design of instructional videos by identifying factors that help or hinder learning, or otherwise impact a learner's experience. This work is informed by research in both visual presence and dialogue domains, and applies frameworks used disparately in each in a single study. To our knowledge, this is the first experiment to do so.

Chapter 2

Background and Related Works

To provide context for the present study's goals and methodology, we now describe related research, including the theoretical foundations that shaped our research questions and hypotheses.

2.1 Monologue presentations and the role of instructor presence

Narrated instructional slideshow presentations typically require little more than a slide deck, microphone, and video editing software to produce, making them relatively resource-efficient options for instructional video producers. But the simplicity of this format can also lead to the same weaknesses associated with traditional lectures: inability to foster critical thinking, deep understanding, and application of knowledge (Bligh, 1998). Unless lectures are concise, interspersed with discussion breaks, or supplemented with other activities like social interaction (e.g., discussion of content), they run the risk of not fostering student engagement (Bligh, 1998). This is an issue because engagement is an important factor for learning (Bates, 2015).

In an effort to increase learner engagement while maintaining a simple, easily-replicable video format, some researchers posit that adding social elements to a presentation takes it a step closer to paralleling a social interaction (Mayer et al., 2003). This reasoning is supported by social agency theory, which postulates that the presence of a social entity or social cues in a learning environment encourages learners to interpret the environment as social communication (Mayer, 2005c; Mayer et al., 2003). Such social cues encourage learners to treat the computer as a partner who has information worthy of attending to and understanding (Mayer et al., 2003). Adding social cues to instructional materials can take the form of using polite instead of direct language (Wang et al., 2008) or conversational wording instead of formal wording (Mayer et al., 2004). Two types of social cue that are most commonly investigated are auditory and visual, as next described.

2.1.1 Auditory presence of the instructor

One of the most basic ways of adding a social element to instructional videos is by including narration provided by the instructor. The instructor's voice is speculated to be a particularly important social component because it primes learners to approach a learning scenario as a human-to-human conversation, thus activating sense-making processes and resulting in better learning outcomes (Mayer et al., 2003). Work by Craig et al. (2002) and Moreno et al. (2001) supports this hypothesis, as learners in both studies who received narrated learning material had higher recall and transfer scores than learners who received text-based material. Benefits of using a human voice with a familiar accent over, for instance, a machine-generated artificial voice are well-supported. Mayer et al. (2003) found that participants presented with learning material narrated by a human voice performed better on recall and transfer tasks than those who heard a mechanical voice, and attributed more positive social characteristics to the human speaker. Those in the human voice condition also rated the material as less difficult to learn, suggesting they experienced lower cognitive load than those in the mechanical voice condition. Atkinson et al. (2005) replicated the learning results of Mayer et al. (2003) with both university and high school students, but did not find any differences in perceived difficulty of content between a human and machine voice. While more detailed analysis on the auditory element is outside the scope of this work, the key takeaway here is that using an instructor's voice to narrate the learning material is an established example of a social element used to enhance learning.

2.1.2 Visual presence of the instructor

Enhancing social presence in instructional videos through visual means is commonly achieved by docking a small video of the instructor (either a human or an animated agent) in a corner of the instructional video frame as they narrate content. In contrast to auditory presence, strong consensus has yet to be reached on the contribution of the instructor's visual presence to instructional videos. Results have been mixed in terms of target outcomes like learning, affect, and cognitive load: while some studies find beneficial effects (Colliot and Jamet, 2018; Kizilcec et al., 2015, 2014; van Gog et al., 2014; Wang and Antonenko, 2017; Wilson et al., 2018), others do not (Berner and Adams, 2004; Homer et al., 2008; Lyons et al., 2012; van Wermeskerken et al., 2018). To help explain these mixed results, we turn to two principles of learning embedded within the Cognitive Theory of Multimedia Learning.

Instructor presence: Seductive detail or social cue?

Mayer's Cognitive Theory of Multimedia Learning (abbreviated to CTML; 2005a) provides guidance for the development of effective educational multimedia by taking into consideration relevant aspects of human cognition. In brief, this theory describes three main assumptions around learning with multimedia: that there are dual channels for processing auditory/verbal and visual/pictorial information, that each channel has limited capacity for processing, and that learning is an active process of attending to, filtering, and integrating new information into existing knowledge. Ideally, educational multimedia would conform with CTML principles to capitalize on learners' auditory and visual processing channels, prevent overloading of stimuli or information, and support encoding of new material. In reality, however, conflicting perspectives of how to best comply with the tenets of CTML create space for ongoing debate.

The visual presence of the instructor in instructional video is a prime example of where two CTML principles provide seemingly contradictory recommendations. For one, CTML posits that instructional material can overextend the cognitive resources of learners by including irrelevant elements. Also referred to as seductive details, these elements, though eye-catching or interesting, do not contribute to given learning objectives (Harp and Mayer, 1998; Mayer, 2005b). The *seductive detail effect* correspondingly refers to when people learn more deeply from material that excludes irrelevant details than from material that includes them (Mayer, 2005b). Given this assumption, the visual presence of the instructor in an instructional video has been hypothesized to be a seductive detail (Colliot and Jamet, 2018; Homer et al., 2008; Moreno et al., 2001). Mayer's image principle (2005c) lends further support to this hypothesis by claiming that evidence for the positive contribution of the instructor's presence to learning is weak. We might therefore hypothesize that the visual presence of the instructor would produce a seductive detail effect and so negatively impact learning by overloading people's cognitive resources.

In contrast to the view that the instructor's presence is a seductive detail, as described above, the second principle of social agency posits that the visual presence of the instructor promotes learning. (Moreno et al., 2001, p. 185) outline this perspective as a "constructivist hypothesis", where the increased social presence conveyed by the instructor promotes interest in learning and motivation to learn the material. For example, Lester et al. (1997) found a positive impact on participants' learning outcomes and perception of the learning experience from the mere presence of a cartoon-like instructor agent. Therefore, under this constructivist hypothesis, we would expect a *social presence effect* to enhance learners' motivation and contribute to better affective and learning outcomes when the instructor is visually present (Dunsworth and Atkinson, 2007; Moreno et al., 2001).

These opposing perspectives contextualize much of the research that has been done on instructor visual presence. While it could be that a video of the instructor is an extraneous element that distracts learners and increases cognitive load, it could also be that the positive effects of social presence nullify impairments to learning by increasing learner motivation. These arguments rely on the notion that video observers actually pay attention to the instructor when he or she is visually present. Eye tracking studies can provide some evidence in this regard since they track learners' gaze patterns, which are generally accepted as an approximation for where cognitive resources are being directed, according to the eye-mind hypothesis (Hyönä, 2010; Just and Carpenter, 1980). Thus, throughout this thesis when we refer to 'attention', we mean visual attention, and by extension, the focus of cognitive resources. We now turn to a review of relevant eye tracking studies.

Eye tracking studies

Eye tracking technology is helpful in exploring instructor visual presence from the seductive detail perspective because it allows researchers to identify where people are visually focusing when watching instructional videos. This adds insight into what elements attract learner attention and for how long or how often. If, for instance, people spend more time looking at the instructor than at the instructional materials on the slide, this could be evidence for the seductive detail effect.

A strong example against the seductive detail effect emerged in the work of van Gog et al. (2014). The authors showed one group of participants a video of an instructor seated at a table solving a physical puzzle. A second group was shown the same video, but the frame was cropped so that the instructor's face was not visible. Both groups were given two trials of video watching and attempted the same physical puzzle as was demonstrated in the video after each trial. There were no significant differences in the amount of fixations each group had on the demonstration area, a noteworthy result considering that participants in the face-visible condition also spent 23% and 17% of fixations in the first and second trials respectively on the instructor's face. This result would suggest that, though about a fifth of participants' fixations were spent looking at the instructor's face instead of the demonstration material, the overall amount of fixations on the learning material was not significantly different than if the instructor's face had not been included. Caution should be used in interpreting this result, however, because of the modest number of participants included in the study's final analysis (n = 18 and 17 respectively for the first and second trials in the face-visible condition and n = 16 and 17 for respective trials in the face-absent condition) and the seemingly large difference between fixations on the demonstration area for each condition (99%, face absent vs. 76%, face present in the first trial; 99%, face absent vs. 82%, face present in the second trial). In terms of learning, the authors posited that the inclusion of the instructor's face may have helped participants learn because they picked up social cues such as the instructor's gaze, which may have flagged and guided their attention. At the same time, the instructor's face did not appear to have a distracting effect on participants' attention to the demonstration — in fact, it promoted learning on the second trial.

The domain selected by van Gog et al. (2014) is rather unique among monologue instructor presence research because a physical puzzle was used and participants were assessed on their ability to solve the puzzle, instead of a more conceptual, non-physical problem in an academic topic. As such, replicating their results in more abstract instructional domains like programming or mathematics would add to the body of work investigating the notion that including the instructor in educational videos is merely a seductive detail. The work of Wang and Antonenko (2017) does offer a contribution in this regard. In one condition, participants in their study viewed material on an easy geometry topic with an inset video of the instructor and a difficult algebra topic without the instructor video. A second condition reversed the presence of the instructor so it was included for the difficult topic only. Overall, participants spent 26% of trial time looking at the instructor in the easy topic, and 22% — significantly less time — in the difficult topic. The authors speculate that this difference may be due to the additional focus needed for the difficult learning content, which drew attention away from the instructor. Dwell time on the learning content was not reported in this work. The recall scores were higher for the easy topic when the instructor video was present than when it was absent. No other differences in recall or transfer scores were significant. Regardless of the topic difficulty, participants also reacted more favourably to material with the instructor present, citing significantly higher levels of satisfaction and perceived learning over material without the instructor. For the difficult topic in particular, participants reported higher mental effort when the instructor was visually absent, but this was not reflected in learning outcomes, as they did not perform significantly worse than counterparts who saw the instructor. Similar to van Gog et al. (2014), Wang and Antonenko (2017) reason that participants in their study may have performed better in the easy topic with the instructor present because the instructor provided non-verbal signals that directed attention to important parts of the presentation. They speculated that this result was not replicated in the difficult topic because the participants needed to expend more cognitive resources on the learning content and thus spent less time attending to the instructor.

As an example of a further probe into the interplay between cognitive load and social presence, Colliot and Jamet (2018) attempted to establish how the visual presence of the instructor affects perceptions of cognitive load, as well as social presence, motivation, interest, and objective learning scores. Two groups of participants were presented with an instructor-present or an instructor-absent instructional video (between subjects). Colliot and Jamet (2018) predicted that, if the instructor was a distracting and irrelevant seductive detail, participants who saw the instructor would perform worse and report a higher level of cognitive load than their counterparts who did not see the instructor. In contrast, if the presence of the instructor contributed beneficial social cues, they expected that learning would be higher in the instructor-present condition, alongside perceived levels of social presence, motivation, and interest. Learning measures were split into three categories, based on information presented orally in the learning material, information presented diagrammatically (visually), and transfer problems. Participants in the instructor-present condition performed significantly better in retention of orally-presented material, but there were no significant differences in the other two learning measures. Participants in the instructor-present condition also spent almost 25% of trial time attending to the instructor and significantly less time attending to the diagram content than their counterparts in the instructor-absent condition. Interestingly, participants who saw the instructor did not report higher levels of cognitive load, nor did they report higher levels of social presence, motivation, or interest. Given the learning results, though, Colliot and Jamet (2018) suggest that their hypotheses around social presence were better supported than their hypotheses around seductive details.

In contrast to the learning gains demonstrated in the aforementioned work, there are also examples of work finding null learning effects. For instance, van Wermeskerken et al. (2018) found no significant differences in recall or transfer scores between participants who viewed learning material with an instructor video and participants who viewed the same material with the instructor removed. Like Wang and Antonenko (2017), the participants in van Wermeskerken et al. (2018) spent a considerable amount of trial time attending to the instructor when he was present (almost 30%) and less time attending to the learning content (52%) — significantly less than the almost 79% of time spent by counterparts who did not see the instructor — but this did not lead to poorer learning outcomes. Kizilcec et al. (2014), in a within-subjects design, presented participants with seven slides that alternated between text plus instructor video and text only. Participants spent an average of 41% of the time looking at the instructor when he was present on the slide, but again, there were no significant learning differences between instructor-present and instructor-absent slides. When considered together, the results of van Wermeskerken et al. (2018) and Kizilcec et al. (2014) do not strongly support either social presence or seductive detail effects when the instructor is included visually because learning was neither negatively nor positively affected. Like the other studies, they do suggest that even though learners split their attention quite noticeably between learning content and instructor, and spend a smaller amount of time looking at learning content than when the instructor is absent, there is no evidence their learning is impaired.

Using eye tracking data, researchers are able to compare participants' attention patterns on videos with the instructor visually present to videos with the instructor visually absent. These comparisons are helpful in supporting or refuting conjectures about the nature of the instructor's presence — namely whether it is harmful or beneficial to learning — based on hypotheses around the seductive detail effect or social presence effect. Evidence gathered thus far suggests that the instructor presence does draw attention away from learning material, but not in a manner that interferes with learning. In fact, in some instances, the instructor's presence has led to enhanced learning.

Studies without eye tracking

Visual presence research that does not involve the use of eye tracking also tends to focus on learning outcomes and psychological measures like cognitive load and social presence. What differentiates these studies from the eye tracking work described above, aside from the obvious lack of eye tracking data, is their use of subjective measures to assess the perceived value of instructor visual presence. For example, some results show that learners respond more favourably to when the instructor is shown, but don't learn more from this condition (Kizilcec et al., 2015; Wilson et al., 2018). In contrast, other work has found negative effects (Homer et al., 2008; Lyons et al., 2012) or neutral effects (Berner and Adams, 2004) with regards to how participants felt or thought about seeing the instructor. We describe these various studies below. As a whole, results from the body of existing research without eye tracking point towards a cautious optimism about the use of instructor video for its positive subjective effects.

For example, in an eight-week field study of students enrolled in a massive open online course (MOOC), Kizilcec et al. (2015, Study 1) allowed students a choice between content with or without a small inset video of the instructor (all participants were shown both options before they could choose). Students could change their choice week-to-week, and watching behaviour was logged for analysis. The authors found that 57% of participants tended to watch the content with the instructor video, 35% tended to watch without the video, and 8% watched both. Subjectively, students who viewed the content with the instructor video reported liking the lectures more, needing to exert less effort, and learning more to a significant degree over those who did not view the instructor. Of the students who responded to an optional feedback survey ($N = 2{,}231$), 59% reported a preference for content with the instructor video and 16% preferred without the video. These results suggest that many students enjoy watching instructional videos with the visual presence of the instructor and also perceive such videos more favourably.

In a competing example, Berner and Adams (2004) split 41 participants from two university classes into two groups per class to watch a slideshow presentation on privacy legislation. The presentation was identical between groups, but a video of the instructor was included on the presentation shown to one group in each class. Participants were then asked to evaluate the lecture and their perceptions of online learning based on the presentation they saw, and finally, completed a knowledge test based on content in the presentation. The authors found no significant effects of group (instructor present, instructor absent) or class (class 1, class 2), nor any significant interaction, on either evaluation of the presentation or content recall, indicating that the addition of the instructor video did not lead to higher satisfaction or learning outcomes. Additionally, while a majority of participants in the instructor-absent condition expressed a desire for the instructor to be included in the presentation, in the instructor-present group, less than half of participants felt that the instructor benefitted their learning, and a third of them found the instructor distracting.

While the work by Kizilcec et al. (2015, Study 1) and Berner and Adams (2004) provide examples of how participants feel about the visual inclusion of the instructor, it is also important to better understand *why* learners might prefer, or not prefer, this format. Some research suggests that learners' individual characteristics have mediating effects on the efficacy and perception of the instructor face. Lyons et al. (2012) compared students' levels of perceived efficacy with the internet and computer-mediated communication — in other words, their technological efficacy — to outcomes of perceived interactivity, learning, and comfort in an online course. Participants in this study were also evaluated over the course of a term, but unlike in Study 1 of Kizilcec et al. (2015), they were assigned to a video condition (instructor-present or instructor-absent) and were unable to change the video format. Lyons et al. (2012) found that including a video of the instructor led to lower levels of perceived learning among students, with technological efficacy being a significant mediator in this result. Specifically, lower perceived learning was more pronounced among students with low technological efficacy. This interaction was also significant in perception of the learning material, where low technological efficacy was associated with perceiving the material as unhelpful when the instructor was present. These outcomes led Lyons et al. (2012) to suggest that students' levels of technological efficacy should be taken into account when developing instructional videos because of how low-efficacious students may perceive the helpfulness and resulting learning of videos. It is important to note that these results should be interpreted with caution because students are notoriously inaccurate at evaluating how much they did (or did not) learn (Anderson and Beal, 1995; Markman, 1979; Pressley et al., 1990); however, this work highlights how individual characteristics can moderate subjective effects of instructor visual presence.

Homer, Plass, and Blake (2008, Study 2) found a similar pattern of results in a classroom setting when investigating the role of visual preference and instructor presence. Participant preferences for information presentation (visual or verbal) were measured alongside their perceptions of social presence and cognitive load after watching a video lecture either with or without the visual presence of the instructor. The authors found no significant differences in cognitive load, social presence, and learning between the two conditions. However, they did find a significant interaction between participants' visual preferences and their cognitive load, where students who had a low preference for visual information reported higher cognitive load when the instructor video was present, and those who had a high preference for visual information reported higher cognitive load when the video was absent. Under slightly different conditions where the instructor was presented at strategic intervals in a video lecture, Kizilcec et al. (2015, Study 2) also found that participants who preferred visual information experienced higher levels of cognitive load when the instructor face was absent.

Taken as a whole, the predominant pattern of results here is exemplified by the recent work of Wilson and colleagues (2018), who found that, when the instructor is present, participants do not learn more, but they do express more positive perceptions. In one experiment, a detriment to learning was associated with the visual presence of the instructor compared to audio-only and audio-plus-text conditions, but this was not replicated in a subsequent experiment (where there was a null effect instead). Moreover, despite not learning more, when shown four different presentation styles, only one of which included the visual presence of the instructor, participants responded most favourably to the instructor-present style, citing higher levels of enjoyment, interest, and perceived learning (Wilson et al., 2018, Experiments 3 & 4).

In sum, research in the domain of instructor visual presence provides moderate support for the inclusion of the instructor in instructional videos. Results from eye tracking studies show that, when present, the instructor draws learner attention away from learning content — to a significant degree, in some cases. However, the distraction does not result in reduced learning; while some studies have found null learning effects and some have found positive effects, none report impaired learning with the presence of the instructor. It has been hypothesized that the non-verbal social cues provided by the instructor serve as guides for learner attention or enhance learner motivation, but more evidence is needed to support this theory since studies have not shown participants to report greater feelings of social presence with the instructor present. Studies that do not use eye tracking, which focus predominantly on subjective measures, have mostly found that learners prefer the presence of the instructor, and cite higher interest and perceived learning when the instructor is present. While learner preference may be an important factor to consider when designing instructional videos, it is also important to understand that other factors, for example, learners' individual characteristics of visual preference and technological efficacy, may mediate how learners perceive such materials.

2.2 Dialogue presentations and the role of vicarious learning

Another vein of research concerned with making online learning materials as effective as possible is rooted in vicarious learning, where students learn through observing and overhearing another student. The theory that people can learn from observing others originates from the work of Bandura et al. (1961) who demonstrated that children, after observing a model behaving aggressively towards a Bobo doll, engaged in similar aggressive behaviour. More broadly, vicarious learning, which is based in social learning theory, proposes that through simply observing a model perform an activity, an observer can learn how to perform the activity (Bandura, 1977). Vicarious learning theory has since been applied to digital education via instructional videos showing a dialogue between a student and a tutor going over the target instructional materials (Chi et al., 2008); these videos are shown to other students with the aim of helping them learn the instructional content. The aim of this form of application is that by watching a dialogue, observers will learn the target material by benefiting from overhearing the student in the video asking questions, receiving explanations, and having misconceptions corrected. A driving factor behind this area of research is exploring how to scale the attractive learning gains of one-on-one tutoring, generally agreed to be the leading gold standard for learning (Bloom, 1984; Cohen et al., 1982). Since one-on-one human tutoring is extremely expensive at a large scale, but videos are not, the hope is that dialogue instructional videos can deliver to larger populations of students the benefits of tutoring without the costs.

To explore the potential benefits of dialogue-based videos, they are commonly compared to the more traditional monologue format that shows an instructor only. It should be noted that, in contrast to the visual presence studies previously discussed, these studies do not focus on principles of CTML as heavily, or at all, because their primary interest is on learning outcomes.

2.2.1 Vicarious learning, monologue vs. dialogue

Several studies suggest that videos that include a dialogue between an instructor and a student (referred to as dialogues below) are more effective for learning than monologue videos that show only an instructor (referred to as monologues below) (Craig et al., 2009; Driscoll et al., 2003; Muldner et al., 2011, 2014). The reasons for this trend have been scrutinized from various angles, including the use of deep questions (Driscoll et al., 2003; Craig et al., 2006), misconceptions (Muller et al., 2007, 2008), and collaboration (Chi et al., 2008, 2017; van Blankenstein et al., 2011). Collectively, these investigations offer evidence that dialogues can improve learning, with results being highly dependent on how the dialogue is formatted and, potentially, perceived by learners.

An early example of work supporting the utility of dialogues over monologues was conducted by Fox Tree (1999), who directed participants to arrange sets of tangrams according to unscripted audio recordings of either a single 'director' giving instructions as a monologue or a dialogue between a 'director' and a 'matcher' who was also trying to order the tangrams according to the director's instructions. Participants heard one monologue and one dialogue each, with order counterbalanced across participants. Fox Tree (1999) found that participants performed better (made fewer positioning errors) after they heard the dialogue. Interestingly, when asked after the activity what they thought about the monologue and dialogue, many participants spoke negatively of the dialogues, citing them as more difficult, despite 70% of those same participants performing equally well or better with the dialogue versus the monologue.

Fox Tree (1999) made a case for dialogue-based instruction using only audio stimuli — but what happens if participants are given visual elements related to a dialogue to look at in addition to the audio? Driscoll et al. (2003) addressed this question by comparing videos of a virtual tutor agent delivering a *monologue* and a virtual tutee and tutor engaged in a question-and-answer style *dialogue*. Using free-response questions as a learning measure, they found that participants recalled more content from the dialogue than the monologue (Driscoll et al., 2003, Experiment 1). Similarly, Muldner et al. (2011) found that learners who observed a video lesson delivered as a dialogue learned more than those who observed a monologue lesson. In this study, participants watched the instructional video either alone or alongside another participant, and the learning gain from the dialogue condition was found to be especially pronounced in participants who viewed the video alone instead of with a partner. Further support for this pattern of results was provided by Muldner et al. (2014, Study 1), who assessed the learning of university student participants after they 1) were tutored one-on-one, 2) watched a dialogue video with a partner, or 3) watched a monologue video with a partner. Again, they found that participants in the dialogue condition learned more than participants in the monologue condition.

The evidence supporting superior learning gains from dialogue videos begs the question of what factors are at play to enhance learning in students watching the videos. Driscoll et al. (2003, Experiment 2) investigated four possibilities, focusing on the contributions to the dialogue of the student (also called the tutee) in the video. They proposed that the tutee in the dialogue improves observer learning because the tutee 1) contributions to the dialogue "[function] as signaling devices similar to headings in printed text", 2) provides repetition of concepts, 3) provides content formed as questions, or 4) contextualizes the information spoken by the tutor (Driscoll et al., 2003, p. 438). These hypotheses were tested by manipulating four presentations with a tutor and tutee: in the first condition, all the tutee's contributions were in the form of deep-level reasoning questions; in the second, the tutee contributions were all shallow-level reasoning questions; in the third, the tutee only made simple comments. As a fourth, 'enhanced monologue' condition, the tutee asked only one broad question at the start of the presentation to justify the tutor delivering the rest of the information as a monologue, including simple comment contributions originally made by the tutee in the third condition. As a contrasting 'monologue-like' condition, a video of the tutor and tutee was shown; similar to the 'enhanced monologue', the tutee only asked a broad question at the start of the presentation and the tutor delivered the rest of the information (without simple comments). Participants were shown four presentations in 'monologue-like' format and four presentations in one of the four dialogue conditions (making discourse format, monologue vs. dialogue, within-subjects and dialogue condition between-subjects). The results revealed that participants in the deep question dialogue condition recalled significantly more relevant details from the dialogue than details from the monologue. No other results were significant across dialogue conditions, suggesting that the shallow questions, simple comments, and enhanced monologue were all similarly effective as the monologue condition. The authors concluded that their results support the hypothesis that deep questions play a role in improved learning, but were unable to address the question of whether or not a monologue condition with similar deep questions would have been equally effective (Driscoll et al., 2003).

Craig et al. (2006) picked up this line of research by specifically probing the effects of deep-level reasoning questions in dialogue and monologue. They first conducted a study with undergraduate students comparing a condition where participants directly interacted with a virtual tutor in an intelligent tutoring system to one of four conditions: participants either observed a monologue showing a virtual tutor going over the instructional materials alone, a dialogue between a human participant and a virtual tutor, a question-and-answer dialogue between virtual tutor and tutee followed by the tutor recapping key concepts in a monologue, or a question-and-answer dialogue between virtual tutor and tutee that included key concepts. The dialogue condition in which deep questions about key concepts were included produced significantly higher learning performance than any other condition (Craig et al., 2006, Experiment 1). No evidence was found that the dialogue conditions were superior to monologue per se; instead, the results seemed to suggest that the presence of deep questions was the important factor. To further investigate, the authors then compared a deep question dialogue condition and a deep question monologue condition to direct interaction with a virtual tutor and to a recording of the direct interaction. They found significantly higher learning outcomes in the conditions involving deep questions — both dialogue and monologue. Between the dialogue and monologue videos that included deep questions, learning outcomes were not significantly different from each other. This again suggests that the questions played a more important role in learning than the format of monologue or dialogue (Craig et al., 2006, Experiment 2). Similar results were later obtained by Gholson et al. (2009) with middle and high school students, further supporting the beneficial effects of deep questions in instructional materials.

Another factor that plays a role in improving learning from dialogues is overhearing misconceptions expressed by the tutee, which helps guide observers to change their own views on a topic. This hypothesis was investigated by Muller et al. (2007). The dialogue condition used in this study included an instructional video with a tutor and tutee, scripted to include information that confronted the tutee's existing conceptions of quantum mechanics. In contrast, the monologue condition showed a tutor who summarized the information without expressing misconceptions. Participants in the dialogue condition had higher learning scores in the post-test than those who viewed the monologue. Muller et al. (2008) later found that misconceptions presented in either dialogue format or monologue format produced better learning outcomes than standard exposition alone, suggesting that regardless of monologue or dialogue presentation, misconceptions enhance learning outcomes.

The research by Craig et al. (2006), Gholson et al. (2009), and Muller et al. (2007, 2008) opens up the possibility that there are not characteristics unique to the dialogue format that enhance learning; when variables such as deep questions and misconceptions are controlled for between conditions, the elevated learning gains related to observing a dialogue format disappear. However, there may be more to dialogue than what is spoken by instructor or student. Muldner et al. (2011) postulate that positive effects of a dialogue format may be due to an increased level of interest among the observers. Such interest may be related to the presence of the student in the dialogue video, with whom participants may relate to and empathize. A dialogue format may also encourage participants to be more engaged (Chi et al., 2008; Muldner et al., 2014). An analysis done by Chi et al. (2017) of data from Muldner et al. (2014) suggests that participants watching a dialogue demonstrated more active and engaged behaviour than participants watching a monologue, and seemed to pay more attention to the student's contributions.

Other mediating factors related to learning gains and video formats — particularly those imposed by participants themselves — are important to consider, though they have not been widely studied. One such factor is whether a video is scripted or naturally occurring, such as a conversation between a tutor and their student in an office hour. Work by Cox et al. (1999) and Cooper et al. (2018) found null results in learning between unscripted monologue and dialogue videos. Cox et al. (1999) speculated that the dialogue condition required participants to expend more cognitive resources than the monologue in order to parse both instructor and student's contributions. As a mediator, however, the dialogue may have held participants' interest more effectively than the monologue, and thus learning was not negatively impacted. Cooper et al. (2018) added to these speculations, finding that a majority of participants, when presented with unscripted monologue and dialogue videos, reported a preference for monologue videos — common reasons given for this preference included the directness of monologue presentation, which contrasted the perceived confusing, disorganized, and misleading nature of the dialogue format. By comparison, participants who preferred the dialogue format focused only on the perceived advantages of the format, such as the inclusion of a student's perspective, the relatability of the tutee, and the tutoring environment of the dialogue (Cooper et al., 2018).

What students prefer in instructional video formats, as demonstrated by the work of Berner and Adams (2004); Cooper et al. (2018); Wilson et al. (2018), is not necessarily reflected in learning outcomes. Nevertheless, the exploration of student preference and other subjective factors is important because when learning outcomes are equivalent, how much a student enjoys a presentation has the potential to contribute to increased self-motivated viewing and decreased attrition rates (Kizilcec et al., 2015). These patterns are notoriously difficult to achieve in online learning settings such as MOOCs, so the idea that formatting videos in ways that most students prefer, or offering students a choice of video format, could be highly beneficial. Further research to develop a clearer understanding of how preference is influenced by video format is therefore essential.

Chapter 3

The Present Study

Studies comparing the effects of instructor visual presence (i.e., work that examines the effects of including or excluding the instructor's image, henceforth referred to as visual presence studies), have focused on how principles of CTML inform the merits and detriments of visually including the instructor — a key *video design* element. Meanwhile, research comparing dialogue videos to monologue videos (i.e., work that includes an instructor and a student they are tutoring; dialogue studies, for brevity), are mostly concerned with how elements of *pedagogical design*, such as deep questions, misconceptions, and collaboration, impact learning. In short, visual presence studies place less emphasis on pedagogical elements and dialogue studies place less emphasis on video design elements. This gap opens up an opportunity for a cross-over of perspectives and analyses to obtain a more holistic view of the effect of the video formats.

The goal of the present study was to address this gap by using CTML principles which, to date, have been the focus of visual presence studies — the seductive detail and social presence effects — to situate the evaluation of dialogue instructional videos alongside monologue videos. For example, questions of cognitive load or social presence induced by the instructor video are often asked in visual presence studies; these questions, however, are also highly relevant to dialogue studies, but to date have not been investigated. Moreover, to our knowledge, though constructs such as cognitive load (Cox et al., 1999), interest (Muldner et al., 2011), and attention (Chi et al., 2017) have been speculated in dialogue studies as contributing to learning results, to date these have not been empirically tested.

The two domains of visual presence and dialogue research guided the present study's design and research questions. To test the effect of instructor and student presence in both monologue and dialogue formats, we developed three types of instructional video that followed a narrated-slideshow convention. These videos formed the basis of our study conditions as follows:

- a *control* condition where the instructor in the video could be heard but not seen
- a *monologue* condition where the instructor in the video could be heard and additionally seen
- a *dialogue* condition where the video included the instructor going over the instructional material with a student (i.e., a tutee)

Several measures, motivated by prior work, were captured to investigate the differences between our study conditions. First, the work of Mayer (2005b) and Ayres and Sweller (2005) suggests that if instructor visual presence is a seductive detail, learners' attention will be drawn to the instructor at the expense of attention to learning content. Additionally, Wang and Antonenko (2017) suggest that learners only dedicate attentional resources to the instructor after they have processed the learning material, implying that participants may pay less attention to the instructor if the learning material requires more cognitive processing. In a dialogue, Chi et al. (2017) posit that learners pay more attention to the student than the instructor. These postulations inspired the use of an eye tracker to capture participants' eye movement data to measure patterns of attention. The first aim of our eye tracking measures was to substantiate the hypotheses of prior work, and the second the aim was to gather data that could potentially explain our other findings. For example, 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 examples in our videos because they were looking at the speaker(s), this would suggest a seductive detail effect.

The second set of measures that we captured were learning outcomes. We were interested in seeing if our instructional videos translated into actual learning outcomes in our participants, and if these outcomes changed differentially by condition, as literature in both visual presence and dialogue domains contain mixed findings. Any conditional effects would then reveal if any particular video format or formats we developed were more or less successful than the others at conveying knowledge.

Five psychological measures were also assessed to explore effects of seductive details and social presence. Cognitive load was selected because a seductive detail effect would predict that higher cognitive load would be registered in conditions with extraneous elements; thus, we wanted to see if visual presence impacted participants' perceived mental effort. We were also interested in whether or not the dialogue format would variably impact cognitive load. Cognitive load is typically seen as having three components: intrinsic load, which describes the inherent difficulty or complexity of the learning material; extraneous load, which describes cognitive load imposed by the design of the learning material (i.e., we would expect this to be higher if learners must parse distracting visuals or difficult-to-read fonts in order to get to the learning content itself); and germane load, which describes cognitive resources needed to process learning materials into long-term storage (Sweller et al., 2011). Because the presence of a seductive detail effect would most likely be driven by the presence of distracting and unnecessary elements, our interest lay in the measure of extraneous load in particular. The second and third psychological variables followed the work of Mayer et al. (2003), who suggest that the addition of social cues to an instructional video contributes to learners' feelings of social presence and interest — thus, measures for both these constructs were included. In the dialogue condition, the measure of social presence was also used to investigate if participants felt differently about the instructor or the student. Finally, measures of mindset and self-efficacy were included as exploratory measures to see if the video conditions differentially affected these constructs. Both measures relate to students' beliefs in their ability: self-efficacy describes how capable one feels about one's ability to accomplish a given level of performance, while mindset refers to how fixed or mutable one believes their intelligence is (Bandura, 1982; Dweck, 1999). A higher level of self-efficacy is indicative of confidence in one's skills and/or abilities. Meanwhile, measures of mindset fall between two main beliefs about intelligence: on one end, an entity (or fixed) mindset views intelligence as a static, inherent trait that cannot be improved. On the other hand, an incremental (growth) mindset views intelligence as malleable. Generally, individuals with a more incremental mindset are more likely to view success as a result of effort and persist in the face of challenges. Prior work suggests that watching a dialogue may improve learner's self efficacy, so we hypothesized that this in turn could influence mindset by making participants more confident in their ability to improve their knowledge and skills (Schunk and Hanson, 1985).

The following three research questions (RQs) framed our work:

RQ1. How does the visual presence (present, absent) and delivery format (monologue, dialogue) in instructional videos impact *visual attention*? Here the main goal is to compare where participants focus attention in the control, monologue, and dialogue videos.

RQ2. What effect does the visual presence and delivery format have on *learning outcomes* in instructional videos? This question aims to explore differences, if any, in learning gains between the control, monologue, and dialogue conditions.

RQ3. How are participant perceptions of *psychological factors* — cognitive load, social presence, interest, mindset of programming, and self-efficacy — influenced by the visual presence in and delivery format of instructional videos? The focus of this question is how these measures differ, at all, between the control, monologue, and dialogue conditions.

We had a number of hypotheses with regards to how the research questions might be answered. With regards to RQ1, we expected the presence of the speakers in the monologue and dialogue conditions to distract participants from the learning content, resulting in less attention paid to the learning content in those conditions (H1a). We also expected the attention to learning content to be further reduced in the dialogue condition because there were three elements to attend to on screen (the instructor face, the student face, and the learning content) (H1b). As an accompaniment to the attention patterns on learning content, we predicted that participant attention would be drawn to the speaker faces, but had no directional hypothesis on whether the speaker in the monologue would draw more attention than the speakers in the dialogue. As a prediction specifically within the dialogue condition, we hypothesized that participants would relate more to the student and would therefore attend more to the student than the instructor (H1c). Because the learning content became progressively more difficult through the experiment, we reasoned that viewing behaviour on the speaker faces would be reduced when harder content was covered (H1d). However, if the faces were a detail that drew learner attention away from the learning content, we expected participants in the monologue and dialogue conditions to spend more time on the review screen to compensate for time lost looking at the speaker(s) (H1e). We thought this effect may be more pronounced in the dialogue condition than the monologue because, again, there was an extra speaker that participants could attend to (H1f).

With regards to RQ2, framing speaker presence as a seductive detail, we expected to see deleterious effects of the speakers' presence on learning, since they distract from participants' ability to focus on the actual learning content (H2a). When comparing monologue to dialogue, the seductive detail effect, and therefore learning impairment, might be particularly strong in the dialogue condition (H2b). In contrast, from the perspective of social agency theory, the speakers' presence could be perceived in a more positive light. Using this lens, we expected that the presence of social cues in the monologue and dialogue conditions would result in higher learning gains in those conditions (H2c). Learning outcomes were expected to be higher in the dialogue condition because the presence of the student would potentially contribute more to socially-motivated learning (H2d).

Finally, with regards to RQ3, we expected that the presence of a seductive detail effect would result in a higher reported extraneous cognitive load for participants in the monologue and dialogue conditions (H3a), with the dialogue condition reporting the highest level of cognitive load because of the additional presence of the student (H3b). An effect of social cues was expected to lead to higher levels of social presence in the monologue and dialogue conditions compared to the control (H3c). Between the monologue and dialogue, a higher degree of social presence was predicted to occur in the dialogue condition (H3d). Within the dialogue condition, an exploratory question we asked was whether or not there were differences in perceived social presence between the instructor and student. Because of the links between social presence and interest, we predicted that patterns of reported interest would follow social presence; namely, that interest would be higher in the monologue and dialogue conditions (H3e) and highest in the dialogue (H3f). Lastly, we hypothesized that the dialogue format, with the student serving as a positive role model, would increase participants' self-efficacy about programming (H3g) and, consequently, their likelihood of having a more incremental mindset (H3h).

This study is the first work to include elements of both visual presence and dialogue research. Here, learning outcomes and psychological measures based on established multimedia learning frameworks aim to give further insight into the effects of the three video formats. Specifically, our comparison of control, monologue, and dialogue conditions effectively informs both visual presence and dialogue domains: any influence of speaker visual presence would reveal itself in similar outcomes in the monologue and dialogue conditions versus the control; meanwhile, any influence of the dialogue format would manifest in similar outcomes in the control and monologue conditions versus the dialogue. Our use of eye tracking further allows us to assess patterns of participant attention when watching instructional videos.

The domain selected for the learning material was programming using the language Python. Programming is an activity that many students find anxiety-provoking and notoriously difficult (Connolly et al., 2009; Jenkins, 2002). Additionally, because various structures, rules, and syntax in programming do not have analogues in everyday life, the likelihood of beginners holding incorrect assumptions is fairly high. For example, variables are assigned values using an equals sign ('=') and may be thought of as containers 'filled' with their assigned value but not directly equivalent to the value. In contrast, people are taught in common mathematics that the equals sign indicates equivalency or interchangeability between items. Directly translating the mental model of an equals sign in math to programming results in errors and thus programming requires the honing of a new mental model. As a potentially intimidating and novel domain for beginners, programming is an ideal topic for learning interventions that can step through the process and address misconceptions. In the case of the present study, the learning material was developed to include these characteristics, as we describe in detail in the following Methods section.

Chapter 4

Methods

4.1 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 enrolled in Psychology majors and were only eligible for participation if they had not previously taken a university-level course in computer programming. Participants who did not give consent did not continue with the experiment and were not included in the final analysis; no participants fit this criteria.

4.2 Materials

4.2.1 Measures

Data were collected from participants in the form of knowledge-based tests, subjective questionnaires, and eye tracking.

```
char = "K"
letter = input("Enter a letter: ")
res = letter + char
print ( res )
```

User types **Z** - what does the program print?

Figure 4.1: Example code tracing question used in the pre-test.

Pre- and post-tests

The 10-question pre-test measuring programming knowledge comprised nine code tracing problems and one programming problem (see Appendix A). The code tracing questions showed a short program and asked participants what the program would output (e.g., see Figure 4.1); the programming problem asked participants to write a program that incorporated a while loop. Eight of the code tracing problems could be answered by directly applying concepts taught in the learning material (constants in the problem descriptions, like variable names, were changed to be different from those given in the learning material). Two problems — one code tracing and one programming — were transfer questions, where participants needed to extrapolate answers from the information given in the learning material in order to answer the question. The post-test followed the same format as the pre-test, with all questions identical in structure but using different values and variables. Each test had a maximum score of 19 points: recall questions accounted for 12 points and transfer questions accounted for the remaining six points.

Self-efficacy

Self-efficacy, used as an indicator of participants' confidence in programming, was measured using items adapted from the Math Confidence Scale developed by Hendy et al. (2014) (Appendix B).

Programming mindset

Mindset has been shown to be applicable not just to general intelligence as originally defined by Dweck (1999), but also to other domains, such as the one used in the present study — programming. Thus, items from Dweck's implicit theory of intelligence scale (Dweck, 1999) were adapted to reflect participant attitudes towards programming (Appendix C).

Interest

Situational interest was measured using a six-item instrument developed and validated by Rotgans and Schmidt (2011) (Appendix D).

Cognitive load

Three forms of cognitive load (intrinsic, extraneous, and germane) were measured using a 10-item questionnaire (Appendix E) developed and validated by Leppink et al. (2013). This instrument was selected because it produces separate scores for intrinsic, extraneous, and germane load, thereby allowing for finer analysis of cognitive load induced by the learning material.[1]

Social presence

Measures of social presence in online learning tend to focus on interactions between learners and instructors, and subsequently, how socially situated learners feel in the online learning environment (Wise et al., 2004; Gunawardena, 1995). While these are important factors to consider for online learning, the context of instructional video-based learning is different because there is no possibility of real interaction

[1] One item, "The activity covered concepts and definitions that I perceived as very complex" was omitted in error

between the learner and the instructor. Thus, we used an instrument from the domain of virtual tutors, specifically, the Attitude Toward Tutoring Agent Scale (ATTAS) devised by Adcock and Van Eck (2005). Originally conceived as a measure of learner affect towards virtual tutors, the scale is adapted in the present study to include five items measuring perceived instructor attitude toward the participant and four items[2] measuring the participant's interest and attention to the instructor (Appendix F). Nine more of the same items were adapted to assess how participants perceived the social presence of the student in the dialogue.

Eye tracking data

Participants' eye movement data were recorded with an eye tracker. The present analysis focuses on fixation data in the form of fixation counts — the number of times a participant's gaze fell on certain areas of interest (AOIs) — and fixation time — the total amount of time a participant spent fixating on a given AOI.

Time spent of review screen

As described below, the instructional materials included two review screens. The amount of time participants spent looking at each screen was extracted from when the content first appeared to when the participant ended the review by triggering a key on the keyboard.

4.2.2 Learning Material

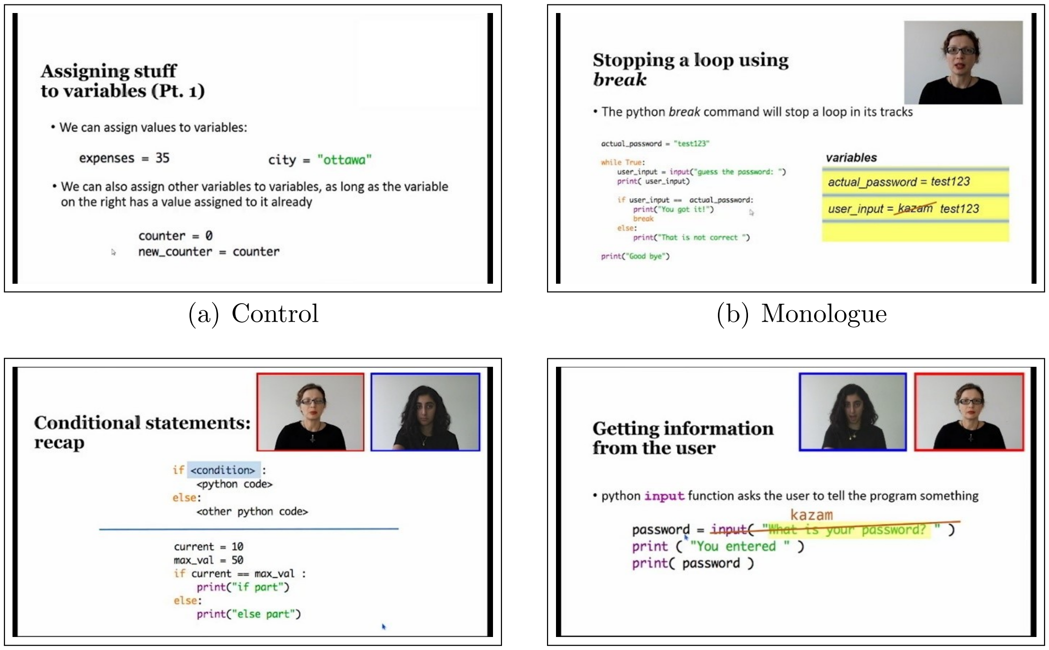
Instructional videos

We created three instructional videos reflecting the study conditions. All three videos included the same instructor talking about the content (the instructor regularly

[2] One item, "The tutor sensed when I needed help", was omitted in error

teaches introductory programming and has a background in the learning sciences). The control video was created by filming the instructor narrating a scripted lesson, but without showing the instructor (Figure 4.2a). The monologue video used the same recording as the control but included the video of the instructor in the top right corner of the screen (Figure 4.2b). The dialogue video featured audio and video of both a student and instructor talking through a scripted lesson and combined with the same slideshow as the control and monologue. The student was a university student who had previously taken a course with the instructor in Python programming. The instructor and student videos were placed side-by-side in the top right corner of the screen. To control for the position of the instructor video, the dialogue condition was split to include a student-instructor configuration, where the student was placed closer to the centre and the instructor in the corner (Figure 4.2c and d). Half the participants in the dialogue condition saw the instructor-student configuration while the other half saw the student-instructor version.

All three videos were scripted based on an a priori-developed Python programming lesson — this was done to control for concept coverage. The three videos included identical content that covered some fundamental programming basics, including variables, conditional statements, loops, and input and output statements. The content was displayed using a "slide" format that is commonly used in instructional materials, with the slides showing examples of Python programs and brief facts about the programs shown (e.g., that a break statement stops a loop, see Figure 4.3). As mentioned above, the videos were narrated with additional information about the content shown on the slide. The narration included questions posed by the instructor and/or student as well as misconceptions that were subsequently refuted — these elements were included based on prior work showing they promote learning (Craig et al., 2006; Gholson et al., 2009; Muller et al., 2008). We controlled for the number of questions



(a) Control (b) Monologue

(c) Dialogue, instructor-student (d) Dialogue, student-instructor

Figure 4.2: Screenshots illustrating the design of the control, monologue, instructor-student dialogue, and student-instructor dialogue videos.

and misconceptions between the three types of videos, with an equal number of questions and misconceptions addressed, but with some content spoken by the student in the dialogue condition (see Appendix G for script excerpt). To control for speaking time in the dialogue condition, the instructor and tutee were given roughly equal speaking time.

The persona of the instructor in all conditions was scripted as that of a personable expert who asked questions and provided feedback on the tutee responses. The tutee was scripted as a focused learner who asked questions and followed the lesson closely, but also demonstrated their role as a programming novice by occasionally verbalizing naïve misconceptions, hesitations, and incorrect answers. In the dialogue script, both the instructor and tutee asked questions to each other and provided explanations. When the tutee gave incorrect answers or expressed misconceptions, they were corrected by the instructor. An adjusted version of the dialogue script was used for the

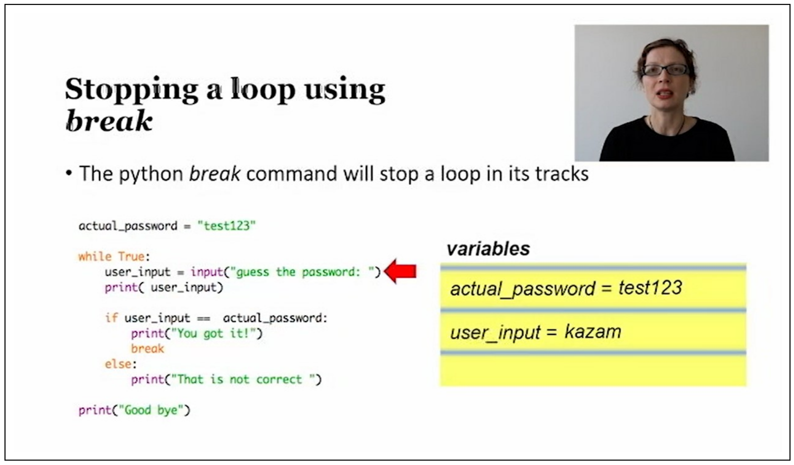


Figure 4.3: Example code tracing question used in the pre-test.

control and monologue videos, where the instructor asked and answered all questions as if addressing the viewer — this approach of instructor question-and-answer is used in prior work using an instructor dialogue (Craig et al., 2006). 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 presentation of the instructional content used in all videos was the same, composed of 17 slides with subtle animations used to highlight important elements or demonstrate progression through code samples (for example, arrows illustrating the act of looping through lines of code). The slideshow component of the videos was created by screen recording a Powerpoint presentation. Cursor movements 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, an 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 to simulate two users (the instructor and student) interacting with the slideshow. Coloured boxes around the instructor and student corresponded to colours of the two cursors present to facilitate attributing each cursor to the correct speaker. The slideshow was recorded twice: once for the control and monologue conditions and once for the dialogue (the latter was needed to account for the addition of the student).

To avoid fatiguing participants by requiring them to watch a long video, each condition's instructional material was split into two parts; participants were given a five minute break in between parts. The first part, covering nine slides, was 10 minutes 14 seconds long in the control and monologue conditions and 10 minutes 16 seconds in the dialogue condition. The second part, covering eight slides, was 16 minutes 58 seconds for the control and monologue and 15 minutes 27 seconds in the dialogue condition.

Videos of the instructor alone and with the student were recorded on a smartphone camera in a classroom setting. Final versions of the instructional material were edited in Adobe Premiere Pro by synchronizing the video and audio from the instructor (and student, for the dialogue video) with the pre-recorded slideshow, and in the case of the control condition, removing the instructor video.

Review screens

Participants were provided a review screen at the end of each video. The screens were still images of single slides shown in the video presentation that covered a key concept from each video (Figure 5.5). Speaker faces were not shown on the review screens.

4.2.3 Apparatus

Eye tracker

A desktop-mounted EyeLink 1000 eye tracker from SR Research Ltd. (Ottawa, Canada) was used to track participant eye movements. Participants were supported

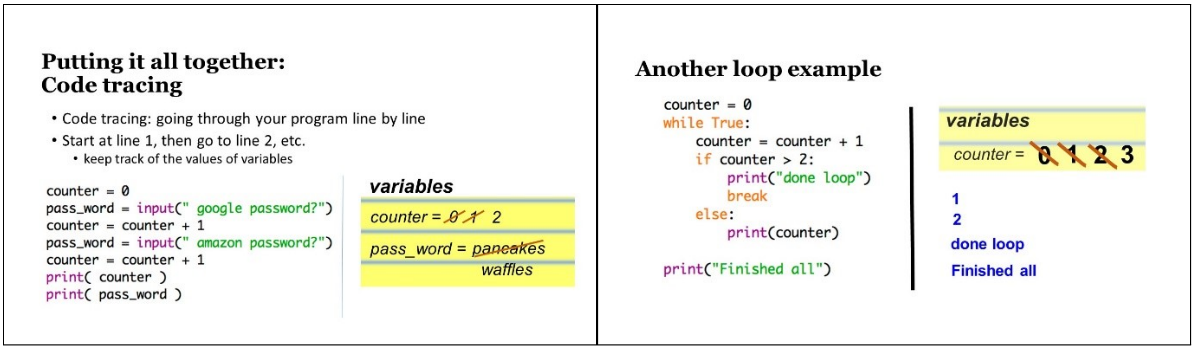


Figure 4.4: Review screens shown at the end of the first video (left) and second video (right).

by a stationary head mount to keep their head still as they watched the learning material, which was displayed on a monitor with a maximum resolution of 1920 x 1080 pixels. The head mount was used based on the advice of SR Research Ltd. that the given EyeLink 1000 model works best with participants stabilized in a mount. Tracking distance from participant to tracker camera was 87 centimetres and camera sampling rate was 2000Hz (.5 milliseconds).

SR Research Experiment Builder

Experiment Builder version 2.1.140 from SR Research Ltd. was used to compose the experimental setup to ensure automatic and consistent stimulus presentation and data collection across participants and conditions. After each video, a still image review screen was programmed to display until a key on the keyboard was pressed, thus allowing participants to view the screen for as long as desired.

4.3 Procedure

Participants were welcomed to the session and invited to sign an informed consent form (Appendix H). Participants who gave consent were given the pen-and-paper pre-test to complete and encouraged to try their best, and to respond with 'I don't know' or 'IDK' to any questions they did not know the answer to. A time limit of 20 minutes was imposed. Following the pre-test, participants filled out online versions of the self-efficacy and mindset questionnaires on a lab desktop computer without a time constraint.

After completion of the pre-test and questionnaires, participants were seated at the eye tracking computer to begin the instructional video portion of the session. Participants were informed that they would be shown two videos on programming, after which their knowledge would be assessed. They were also briefed on the short break in between the videos and the chance to review material using the slides at the end of each video. When comfortably positioned in the head rest, participants were calibrated on the eye tracker. Calibration was completed with the participants' left eye in all cases except four, where refraction from prescription glasses impacted the tracker's accuracy in the left eye, and thus the right eye was used. Assignment to conditions was done in a round-robin fashion (dialogue conditions alternated between instructor-student and student-instructor configuration). While watching the videos, participants were not able to pause, rewind, or fast-forward the material; the videos played in a linear fashion and the corresponding review screen was displayed at the end of each video automatically.

For each review screen at the end of the videos, participants were given a keyboard to terminate the screen when they felt ready to proceed and were not required to remain in the headrest during this time. During the break, participants were allowed to move away from the headrest for the five minute period, and were resituated at the headrest to begin the second video. A drift correct was presented before the start of the second video to ensure participants' initial calibration was still valid; if a participant failed the drift correct, or the researcher observed that tracking had not been accurate during the first video, the participant was recalibrated before proceeding.

Once participants finished viewing the review screen of the second video, they were administered a pen-and-paper post-test, again with a maximum time allotment of 20 minutes, and online self-efficacy, mindset, interest, cognitive load, and social presence questionnaires with no time limit. Finally, participants were thanked for their participation and offered a digital copy of the consent form for their records.

Chapter 5

Results

All results were calculated using IBM SPSS Statistics Version 25. Alpha levels were set at .05 but we considered findings *marginally significant* if p-values were greater than .05 but smaller than .1. Significant and marginally significant results are reported in this section; non-significant results are only reported where they contribute to research questions or hypotheses. Where appropriate, the data was analyzed using ANOVA, and significant or marginally significant results were followed up with Fisher's LSD tests. The use of Fisher's LSD as a post hoc test is debated by some academics because it does not control for alpha; however, our choice to use it was based on whether or not the repeated measures ANOVAs were significant, which would have provided some alpha-level control. Three sets of data will be reviewed in turn: eye tracking results, learning outcomes, and subjective measures.

5.1 Eye Tracking Results

Eye tracking data was extracted for 68 participants (data from the remaining nine participants was unusable due to technical failures; control $n = 21$, monologue $n = 24$, dialogue $n = 23$). To analyse participants' attention to the instructional materials, we

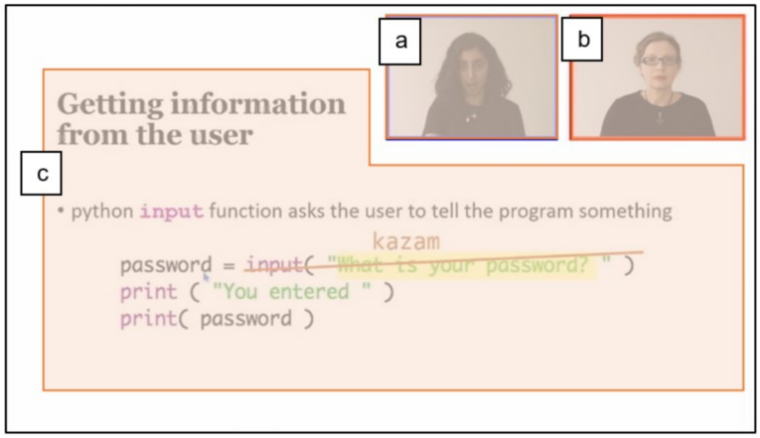


Figure 5.1: AOIs shown in orange for the student (a), instructor (b), and Python content (c). AOIs were not visible to participants.

created areas of interest (AOIs) in the eye tracking data as follows. For the dialogue condition, we created one AOI around the instructor and one around the student, as shown in Figure 5.1 — these are referred to as *instructor AOI* and *student AOI*, respectively. The two AOIs include the whole frame that each speaker was situated in rather than only the face so that fixations that fell just outside the face would still be included. For the monologue condition, the instructor video likewise had a single AOI that included the whole frame, also referred to as *instructor AOI*. In all conditions, the *Python content AOI* encapsulated the title and Python instructional text presented on the slides. Originally created as two separate AOIs for the title and instructional text, the Python content AOI became a consolidation of the two because independent analysis of those areas was ultimately more granular than necessary, since both contained instructional text. The Python content AOI was exactly the same size and in the same place in each condition. As such, the dialogue condition contained AOIs for the instructor, student, and Python content, the monologue condition contained AOIs for the instructor and Python content, and the control condition contained only one AOI — that of the Python content.

The two videos that each participant watched were analysed as separate trials, with the first video henceforth referred to as *Trial 1* and the second video as *Trial 2*. This distinction was made because the content in the second video, covering conditional statements and loops, was more complicated than that in the first video, which addressed simpler, more fundamental concepts like variables and print functions. Keeping trial data separate allowed for analysis and comparison of each video's content, which did in fact have differing effects, as described in the following subsections.

We used two measures extracted from the eye tracking data to operationalize attention to the instructional materials. The first, dwell time, is the time that a participant spent fixating on an area of interest. The second, fixation count, is the number of fixations a participant had on an area of interest. As previously described, the videos were of slightly different lengths (control and monologue videos were the same length, but dialogue videos were just slightly longer; additionally, the second video in each condition was longer than the first video — see *Instructional Videos* section in Methods). Thus, using dwell time and fixation count measures outright might give the false impression that the longer videos held more attention, when the video length could be the factor accounting for a larger measure. To control for this, dwell time was divided by video (trial) length and fixation count was divided by total video (trial) fixation count for each AOI. These transformed measures, dwell time percentage and fixation percentage, will be referred to respectively as *dwell time %* and *fixation %* for brevity. Note that the denominator for dwell time % was total *trial* time, so summing the dwell time % from all AOIs does not necessarily equal 100% because instances when participants looked away from the AOIs or were not tracked by the camera were not included.

The following eye tracking results address RQ1 and its respective hypotheses,

Table 5.1: *Means and standard deviations for Python content dwell time % and fixation % between conditions*

		Dwell time percentage		Fixation percentage	
	Trial	*M*	*SD*	*M*	*SD*
Control	1	71.80	16.99	96.70	2.69
	2	70.15	15.93	96.94	2.37
	Total	70.97	15.92	96.82	2.23
Monologue	1	57.95	9.04	81.63	6.67
	2	57.50	11.13	81.56	7.88
	Total	57.72	9.59	81.60	6.97
Dialogue	1	52.49	9.23	76.32	5.59
	2	55.95	11.61	79.31	6.36
	Total	54.22	10.02	77.82	5.55

which probe how visual attention behaviour differed between conditions.

5.1.1 Attention to Python content

With our first hypothesis (H1a), we predicted that less attention would be paid to the Python content in the monologue and dialogue conditions because of the presence of the speaker faces. We also predicted this reduced attention to be especially prominent in the dialogue condition (H1b). To check these hypotheses, we compared participant attention to the Python content AOI.

The descriptive statistics for all conditions are presented in Table 5.1. A brief note should be made about fixation time unaccounted for by Python content in the control condition (e.g., almost 30%). While some of this may be attributed to lost samples on the part of the eye tracker, it is most likely that participants simply did not look

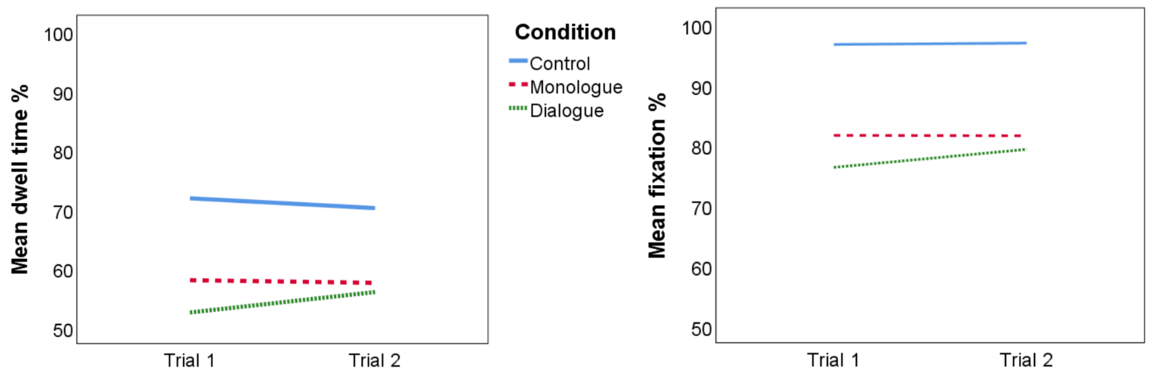


Figure 5.2: Mean attention to Python content from Trial 1 to Trial 2 across conditions, for dwell time % (left) and fixation % (right).

at the screen for portions of the videos. The nature of the Python content is thought to have played a role, as we will later discuss. Returning to the attention data that we did capture, as predicted, participants in the monologue and dialogue conditions had markedly lower dwell time % and fixation % than participants in the control condition. This was expected because the speakers in the monologue and dialogue videos were elements that drew participants' attention away from the Python content; meanwhile, participants in the control condition had no other elements to look at besides the Python content. The difference between the means of the monologue and dialogue conditions was less pronounced, but the dialogue participants appeared to be slightly less attentive to the Python content than their monologue peers.

Our interpretations were confirmed with a repeated measures analysis of variance (ANOVA) for each dependent variable (dwell time %, fixation %) with *trial* (first, second) as the within-subjects variable and *condition* (control, monologue, dialogue) as the between-subjects variable. There was a significant main effect of *condition* for dwell time % on the Python content, $F(2, 65) = 11.76$, $p < .001$, $\eta_p^2 = .27$, and for fixation %, $F(2, 65) = 75.47$, $p < .001$, $\eta_p^2 = .70$. Planned comparisons using Fisher's LSD revealed that participants in the control condition spent a significantly higher percentage of trial time dwelling on the Python content than participants in the monologue condition (dwell time %: $t(43) = 3.69$, $p < .001$, $d = 1.01$; fixation %: $t(43) = 9.44$, $p < .001$, $d = 2.94$) and dialogue condition (dwell time %: $t(42) = 4.62$, $p < .001$, $d = 1.26$; fixation %: $t(42) = 11.66$, $p < .001$, $d = 1.26$). Participants in the monologue condition also had a significantly higher fixation % than their counterparts in the dialogue condition, $t(45) = 2.40$, $p = .02$, $d = .36$. The difference in dwell time % between the monologue and dialogue conditions was not significant.[1]

There was also a significant interaction between *condition* and *trial* on dwell time % and fixation % (dwell time %: $F(2, 65) = 11.76$, $p < .001$, $\eta_p^2 = .09$; fixation %: $F(2, 65) = 4.31$, $p = .02$, $\eta_p^2 = .12$). As shown in Figure 5.2, participants in the dialogue condition spent a larger percentage of trial time and fixations in Trial 2 looking at the Python content than in Trial 1. Counterparts in the control and monologue conditions did not parallel this viewing behaviour and instead spent comparable amounts of dwell time and fixation % on the Python content between trials.

Given the above findings, our first hypothesis, H1a, was confirmed: participants in the monologue and dialogue conditions paid less attention to the Python content than participants in the control condition. Our next hypothesis H1b, which predicted that dialogue participants would pay the least attention, was supported in fixation %; additionally, the interaction between *condition* and *trial* showed that dialogue participants actually spent more time and fixations on the Python content, but only in Trial 2.

[1] For the sake of completeness, the planned comparisons were re-run using Tukey's test, which is more conservative, and the pattern of results was the same.

Table 5.2: *Means and standard deviations for dwell time % and fixation % on the speakers in monologue and dialogue conditions*

		Dwell time percentage		Fixation percentage	
	Trial	*M*	*SD*	*M*	*SD*
Monologue	1	19.61	6.65	15.75	5.74
	2	17.76	9.05	14.26	6.04
	Total	18.69	7.66	15.00	5.45
Dialogue	1	25.52	9.41	20.50	5.78
	2	21.28	8.16	17.94	5.79
	Total	23.40	1.63	19.22	5.58
Total	1	22.50	8.57	18.07	6.18
	2	19.48	8.72	16.06	6.14

5.1.2 Attention to the speakers

Another facet of RQ1 had to do with patterns of attention to the speakers in the monologue and dialogue conditions. While we did not hypothesize about differences in attention to the speakers between the conditions, within the dialogue condition, we predicted that more attention would go to the student than the instructor (H1c) and that across conditions, attention to the speakers would be reduced when more difficult content was covered (i.e., an effect of trial; H1d).

The descriptives for monologue and dialogue conditions are presented in Table 5.2. In order to compare these two conditions on a single variable, data from the instructor and student in the dialogue condition were combined to form one measure of speaker presence, and this was compared to the sole instructor presence in the monologue condition. Two repeated measures ANOVA then were conducted for attention to the speaker(s) with *trial* (first, second) as the within-subjects variable and *condition*

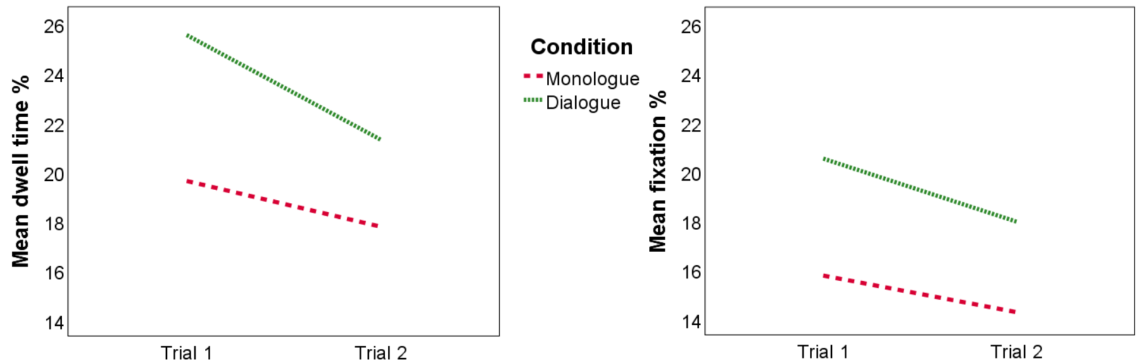


Figure 5.3: Mean attention to the speaker(s) in each trial between conditions for dwell time % (left) and fixation % (right).

(monologue, dialogue) as the between-subjects variable, one for dwell time % and one for fixation %. As shown in Table 5.2, participants in the dialogue condition had a higher average dwell time and fixation % on the speakers than their counterparts in the monologue condition in both trials. These trends were subsequently confirmed by significant differences between conditions, $F(1, 45) = 4.27$, $p = .045$, $\eta_p^2 = .08$ for dwell time % and $F(1, 45) = 7.02$, $p = .01$, $\eta_p^2 = .14$ for fixation %. This is an interesting result given that the dialogue video was scripted so that the combined amount of speaking time that the student and instructor had would be roughly equivalent to the speaking time of the instructor in the monologue. However, dialogue participants may have paid more attention to the speakers in their condition overall simply because there were two people to attend to.

Hypothesis H1d was confirmed for the speakers with a main effect of *trial*, where collapsed across conditions, participants spent a greater proportion of attention on the speakers in Trial 1 than in Trial 2, $F(1, 45) = 12.18$, $p < .001$, $\eta_p^2 = .21$ for dwell time % and $F(1, 45) = 11.19$, $p = .002$, $\eta_p^2 = .20$ for fixation %. There was no significant interaction between *trial* and *condition*, meaning that the reduction in attention from Trial 1 to Trial 2 was similar in both conditions (see Figure 5.3).

Attention to instructor vs. student in the dialogue condition

Hypothesis H1c followed the work of Chi et al. (2017), who proposed that observers pay more attention to the student than the instructor in a dialogue presentation. Before we verified if this conjecture held in our data, we checked if the way the instructor and student were positioned in the video affected outcomes. Recall that half of participants in the dialogue condition were placed in an instructor-student order ($n = 11$), where the instructor's video was closer to the centre of the screen, and the other half of participants were placed in a reverse, student-instructor order ($n = 12$, refer to Figure 4.2c and d).

To identify any order effects, mean dwell time % and fixation % on the speakers, shown in Table 5.3, were compared between dialogue orders. While there was not an extreme difference between the conditions, participants in the instructor-student order had a slightly smaller average dwell time % on the student than participants in the student-instructor order. To check if this effect was reliable, a repeated measures ANOVA was conducted for each dependent variable (dwell time %, fixation %) with *speaker* (instructor vs. student) and *trial* (first vs. second) as the within-subjects variables and *dialogue order* (instructor-student vs. student-instructor) as the between-subjects variable.

There was a marginal interaction between *speaker* and *dialogue order* for fixation %, $F(1, 21) = 3.71$, $p = .07$, $\eta_p^2 = .15$, where participants in the student-instructor order looked at the student slightly more than participants in the instructor-student order. This interaction was not significant for dwell time %, $F(1, 21) = .53$, $p = .47$, $\eta_p^2 = .03$. This can be interpreted as participants in the student-instructor order fixating slightly more on the student than the instructor than participants in the instructor-student order, but not so much that the dwell time % was significantly higher on the student. Since there was no evidence that the

Table 5.3: *Means and standard deviations for dwell time % and fixation % on the* instructor AOI *and* student AOI *in the dialogue condition*

		Dwell time percentage		Fixation percentage	
	Dialogue order	*M*	*SD*	*M*	*SD*
Instructor	Instructor-student	12.31	4.86	10.60	3.35
	Student-Instructor	12.62	4.93	10.58	3.37
	Total	12.47	4.79	10.59	3.28
Student	Instructor-student	10.14	4.29	7.31	2.54
	Student-instructor	11.65	4.82	9.85	3.02
	Total	10.90	4.54	8.58	3.03

presentation order of the speakers affected viewing behaviour to a significant degree, the two dialogue orders were collapsed to form a single dialogue condition.

To now answer H1c, dialogue participants had a significantly higher fixation % on the instructor, $F(1, 21) = 9.17$, $p = .006$, $\eta_p^2 = .30$. There was also a marginal main effect of *speaker* on dwell time %, $F(1, 21) = 3.63$, $p = .07$, $\eta_p^2 = .15$, with participants spending a higher average proportion of trial time looking at the instructor than the student (See Figure 5.4). Overall, we can conclude that participants did not pay more attention to the student versus the instructor, and in fact, tended to look more at the instructor; as such, hypothesis H1c was not supported.

5.1.3 Time spent on the review screen

As an accompaniment to the Python content and speaker eye data analysis, we examined the time spent on the review screen time in each condition, motivated by hypotheses H1e and H1f. Here, participants in the speaker presence conditions, and particularly the dialogue condition, were predicted to spend longer on the review screen to make up for reduced time on Python content during the videos. Absolute

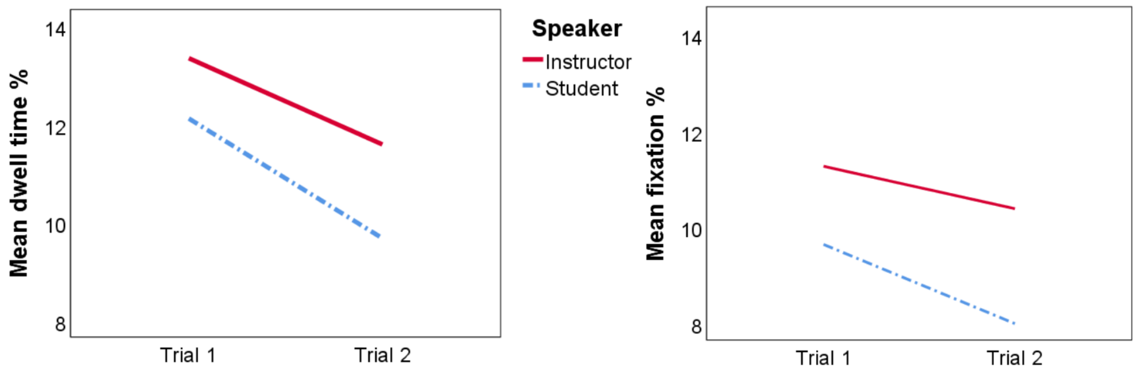


Figure 5.4: Mean attention to the speaker(s) in each trial between conditions for dwell time % (left) and fixation % (right).

times were used in this analysis instead of dwell time percentages because the review screen was not embedded in the instructional videos in the way that, for example, the Python content AOI was. Additionally, because participants were not required to stay in the head mount during the review portion and could end the review whenever they wished, the total elapsed time that the review screen was present is considered its dwell time, and fixations were not included.

The time that participants spent on the review screen is shown in Table 5.4. Participants in the monologue condition spent the highest mean time on the review screen, followed by the control and dialogue conditions (illustrated in Figure 5.5). A repeated measures ANOVA was conducted on *review screen time* as the dependent variable, *trial* (first vs. second) as the within subjects factor, and *condition* (control vs. monologue vs. dialogue) as the between subjects factor. While the review times of all conditions were not significantly different from each other, $F(2, 65) = 1.05$, $p = .36$, $\eta_p^2 = .03$, there was a significant main effect of trial, $F(1, 65) = 4.16$, $p = .046$, $\eta_p^2 = .06$. This indicates that collapsed across condition, participants spent more time on the review screen in Trial 1 than Trial 2. The interaction between *trial* and *condition* was not significant, meaning that time spent reduced in Trial 2 by a

Table 5.4: *Means and standard deviations for time spent on the review screen for participants in each condition, shown in seconds*

	Trial	M	SD
Control	1	25.99	17.35
	2	24.77	16.05
	Total	25.38	14.90
Monologue	1	32.83	17.35
	2	28.15	16.06
	Total	30.49	14.90
Dialogue	1	27.43	17.35
	2	27.43	17.35
	Total	22.04	16.06
Total	1	28.75	17.38
	2	24.99	16.09

comparable amount in all three conditions.

As a result of these findings, we have no support for H1e and H1f since participants in all conditions spent similar amounts of time on the review screen, with no pronounced differences between the monologue and dialogue conditions.

5.2 Learning Outcomes

Our second research question (RQ2) and related hypotheses were concerned with learning outcomes. To reiterate, we were interested in whether or not effects of seductive details or social presence would be detectable through learning outcomes (H2a and H2b predicted lower learning outcomes, particularly in the dialogue condition; H2c and H2d predicted higher, particularly for dialogue). Our hypotheses were motivated by prior work on visual presence and dialogue format that have shown mixed

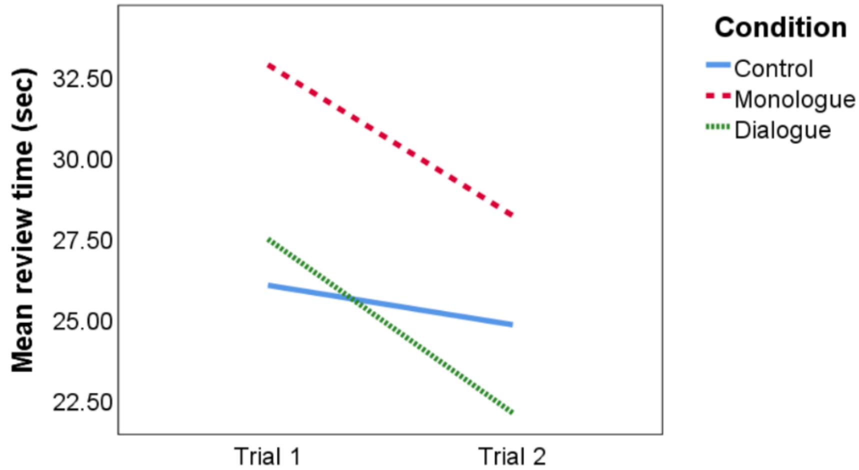


Figure 5.5: Review screen time for each trial between conditions.

results on the positive impacts of their respective instructional video manipulations.

Learning was measured using the standard method of calculating gains from pre-test to post-test; the maximum achievable score on each test was 19 points. The following analysis is based on data from all 77 participants (control $n = 26$; monologue $n = 26$; dialogue $n = 25$).[2] Participant pre-test and post-test scores are presented in Table 5.5. Prior to our main analysis, we confirmed that all participants were starting with equivalent knowledge before the intervention based on their pre-test scores. First, an independent samples t-test was run with the two dialogue orders to determine their equivalence. Since no significant difference was found, $t(23) = 1.06$, $p = .30$, the two dialogue orders were collapsed into a single dialogue condition. Next, a one-way ANOVA was conducted between all three conditions' pre-test scores and again, there was no significant difference, $F(2, 74) = 1.13$, $p = .33$, indicating all participants' programming knowledge was comparable before they watched the learning materials.

[2]We also ran analysis of learning data excluding participants who were at ceiling at pretest ($< 90\%$) or who did not gain from pre-test to post-test, which were potential indicators of a lack of effort in watching the learning materials or in taking the post-test. Three participants were removed in this way from the monologue condition and one from the dialogue. Results were the same as

Table 5.5: *Means, standard deviations, and gain scores for learning in each condition*

	Pre-test (/19)		Post-test (/19)		Gain	
	M	*SD*	*M*	*SD*	*M*	*SD*
Control	3.69	3.52	10.54	4.40	6.85	4.33
Monologue	3.85	3.36	10.96	4.29	7.12	4.08
Dialogue	2.54	3.26	10.42	4.64	7.88	4.34

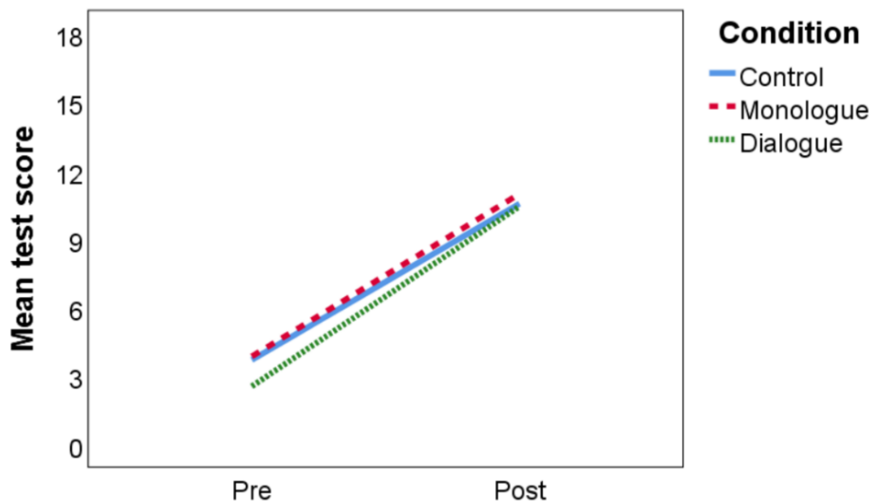


Figure 5.6: Changes in test scores from pre- to post-test between conditions.

Participants in the dialogue condition achieved the highest gains from pre-test to post-test, followed by monologue participants and control participants. These effects, however, were not significant: a repeated measures ANOVA with *condition* as the between subjects factor and *time point* (pre- vs. post-test) as the within subjects factor did not find a significant interaction between *condition* and *time point*, $F(2, 74) = .38$, $p = .69$, $\eta_p^2 = .01$. In other words, increases in learning scores were not affected by any given condition (see Figure 5.6).[3]

One remaining consideration of our analysis was that our test had two types of

when these participants were included.

[3]Some researchers advocate for the use of ANCOVA; we re-ran this analysis with pre-test score as the covariate and received a similar null result, $F(2, 73) = .09$, $p = .91$, $\eta_p^2 = .003$

questions: *recall* questions that tested content similar to the instructional material, and *transfer* questions that required learners to use their knowledge in novel applications. Previous studies have investigated these learning types separately (Homer et al., 2008; Muldner et al., 2011), so we followed suit to see if any differences occurred on this more granular level. Two repeated measures ANOVA for each question type (one for recall, one for transfer) were conducted using *time point* (pre- vs. post-test) as the within subjects variables and *condition* between subjects. Neither test found a significant interaction between *time point* and *condition* (recall: $F(2, 74) = .32$, $p = .73$, $\eta_p^2 = .008$; transfer: $F(2, 74) = .34$, $p = .72$, $\eta_p^2 = .009$). These results indicated that gains in each question type were not different across the video conditions.

Thus, overall we did not find that condition had an effect on learning outcomes (refuting hypotheses H2a, H2b, H2c, and H2d). However, we did find a main effect of *time point*, $F(1, 70) = 257.75$, $p < .001$, $\eta_p^2 = .79$, meaning that, collapsed by condition, participants did learn from the intervention because post-test scores were significantly higher than pre-test scores.

5.3 Questionnaire Results

Subjective questionnaires aimed to measure effects of the control, monologue, and dialogue conditions on mindset, self-efficacy, cognitive load, interest, and social presence, according to RQ3. We hypothesized that participants would perceive higher levels of extraneous cognitive load with the visual presence of the speakers, especially in the dialogue condition (H3a, H3b). However, we were also interested in the positive effects of visual presence on perceived social presence (H3c), especially in the dialogue condition (H3d), as well as on interest (H3e, H3f). Lastly, we predicted that the dialogue condition would influence participants' self-efficacy (H3g) and thus their

Table 5.6: *Means and standard deviations for the three levels of cognitive load between conditions*

	Intrinsic load		Extraneous load		Germane load	
	M	*SD*	*M*	*SD*	*M*	*SD*
Control	3.81	2.04	1.37	1.39	7.01	1.69
Monologue	3.50	2.95	1.76	1.74	7.51	2.30
Dialogue	3.02	1.78	1.75	1.79	7.26	2.02

mindset on programming would become more incremental (H3g).

5.3.1 Cognitive load

As shown in Table 5.6, the monologue and dialogue condition participants had higher levels of extraneous load and lower levels of intrinsic load than the control condition. However, these differences were not significant, as verified with a repeated measures ANOVA using the three levels of cognitive load (intrinsic, extraneous, germane) as the within-subjects variable and condition as between-subjects. The sphericity assumption was violated, so the Greenhouse-Geisser correction was used, $F(3.68, 128.83) = 1.03$, $p = .39$, $\eta_p^2 = .03$. Our hypotheses predicting that participants in the monologue and dialogue conditions would experience higher levels of extraneous cognitive load were therefore unsupported (H3a and H3b). The two other measures of cognitive load, intrinsic and germane, were also not significantly different between conditions.

As a more qualitative observation, the reported levels of intrinsic load fell towards the low end of the Likert scale (out of 10) and exhibited a positive skew (.71). Given that programming was selected as the learning topic for its challenging nature, the fact that participants did not feel particularly strained by the complexity of the learning materials is somewhat surprising.

5.3.2 Social presence and interest

The visual presence of the speakers in the monologue and dialogue conditions were predicted to yield higher participant perceptions of the speakers' social presence (H3c, H3d), and by extension, their interest in the topic (H3e, H3f), over the control condition. Descriptives are presented in Table 5.7 (for concision, 'instructor' is abbreviated to 'Inst.', 'student' to 'Stud.', and 'participant' to 'Part.'). Social presence comprised two components: how positive participants felt the speaker(s') attitude was/were towards them, and how interested and attentive the participants felt towards the speaker(s). When comparing between all three conditions, only the measures of social presence for the instructor were used, since the dialogue condition was the only one that had the student. Comparing means between conditions, the social presence measures from dialogue participants were slightly higher than the other two conditions, and slightly lower in the monologue (see Figure 5.7). To investigate further, a one-way ANOVA was run for each social presence component (perceived instructor attention towards participant, participant attention and interest towards the instructor) with *condition* as the between subjects factor. There was no significant difference between conditions on how attentive participants felt towards the instructor, $F(2, 76) = 1.20$, $p = .31$, $\eta^2 = .03$, and there was a marginally significant difference between conditions on the instructor's attitude towards participants, $F(2, 76) = 2.61$, $p = .08$, $\eta^2 = .07$. Post-hoc tests revealed that only the difference between the monologue and dialogue conditions was significant, $p = .03$, where dialogue participants rated the instructor's attitude more favourably than monologue participants.

Similarly, a one-way ANOVA for situational interest between conditions found no significant differences in how interested participants felt towards programming after the intervention, $F(2, 74) = .02$, $p = .98$, $\eta^2 < .001$. These findings failed to support the hypotheses we generated predicting that perceived social presence would

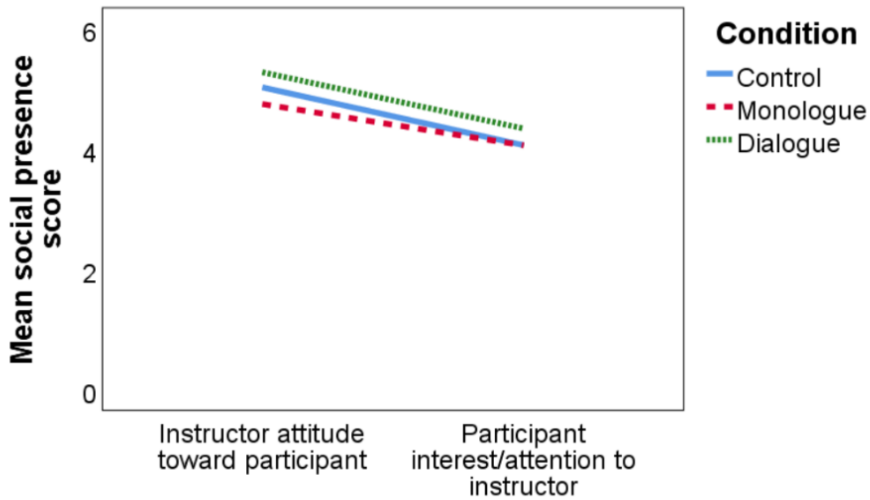


Figure 5.7: Mean scores for social presence components between conditions.

be higher in the speaker presence conditions (H3c and H3d), or that interest would also be higher in those conditions (H3e and H3f).

Based on suggestions from previous work that participants are more likely to be interested in the student in the video (Chi et al., 2017; Muldner et al., 2011), we compared social presence scores of the instructor and the student in the dialogue condition. Two paired samples t-tests were conducted, one for instructor vs. student attitude towards the participant and one for participant interest/attention to the instructor vs. student. We found no significant difference in how interested/attentive participants felt towards the instructor versus the student, $t(24) = .99$, $p = .33$, $d = .20$. However, there was a significant difference in perceptions of the instructor's attitude over perceptions of the student's attitude, $t(23) = 12.21$, $p < .001$, $d = 2.39$, where participants perceived the instructor to be more attentive towards them than the student was.

Table 5.7: *Means and standard deviations for social presence and interest measures between conditions*

	Social Presence								Interest	
	Inst. attitude		Part. interest/ attention to inst.		Stud. attitude		Part. interest/ attention to stud.			
	M	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
Control	5.02	.79	4.06	1.05	-	-	-	-	2.92	.78
Monologue	4.75	.80	3.87	1.08	-	-	-	-	2.97	.78
Dialogue	5.23	.65	4.33	1.11	4.33	.57	4.16	1.26	2.93	.97

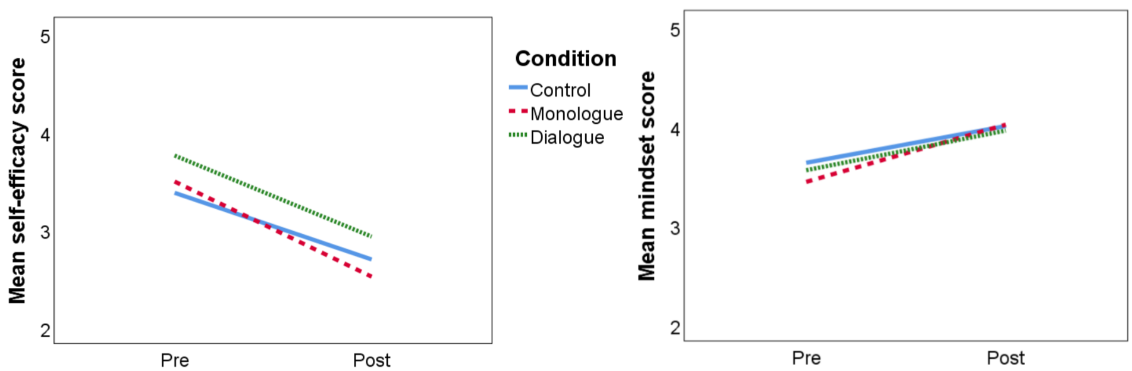


Figure 5.8: Mean self-efficacy (left) and mindset (right) scores before and after interventions for each condition.

5.3.3 Self-efficacy and mindset

The final two psychological measures in the present study were self-efficacy and mindset. Given that self-efficacy has been shown in prior work to improve after a learner watches a dialogue video, we hypothesized that the dialogue condition would bestow a particular boost to this measure (H2j, H2k). Lastly, we predicted that speaker visual presence, in the dialogue format in particular, would influence participant mindset towards a more incremental outlook (H2h, H21).

Descriptives are presented in Table 5.8. Repeated-measures ANOVAs for each

Table 5.8: *Means and standard deviations for self-efficacy and mindset measures between conditions*

	Self-efficacy				Mindset			
	Pre		Post		Pre		Post	
	M	SD	M	SD	M	SD	M	SD
Control	3.37	.68	2.69	.74	3.63	.82	4.00	.83
Monologue	3.49	1.05	2.52	.97	3.44	.74	4.01	.77
Dialogue	3.75	.85	2.93	1.07	3.56	.47	3.96	.72

dependent variable (self-efficacy, mindset) were run with *time point* (pre- vs. post-intervention) as the within subjects factor and *condition* as the between subjects factor. Mean self-efficacy scores were notably lower in each condition post-intervention, with the monologue condition presenting the lowest score (see Figure 5.8). However, the interaction between *condition* and *time point* was non-significant, $F(2, 74) = .85$, $p = .43$, $\eta_p^2 = .02$, indicating that the self-efficacy of participants in any given condition did not change significantly from pre- to post-intervention. A somewhat reverse trend was present in mindset scores, where we observed a general increase in scores from pre- to post-intervention, though the gain was rather uniform (about .40 points) across conditions. Again, there was no significant interaction between *condition* and *time point* to indicate that the change in mindset in any given condition was more pronounced than the others, $F(2, 74) = .50$, $p = .61$, $\eta_p^2 = .01$.

Though our predictions around the positive influence of dialogue on self-efficacy and mindset (H3g, H3h) were unsupported, when collapsed across condition, the change in self-efficacy score after the interventions was significant, $F(1, 74) = 81.65$, $p < .001$, $\eta_p^2 = .53$. This indicated that participants as a whole felt less confident after the intervention. The effect of *time point* was also significant for mindset scores, $F(1, 70) = 23.89$, $p < .001$, $\eta_p^2 = .24$, though in the reverse direction — the mean mindset score of participants increased after the intervention. This indicates that participants moved towards a more incremental view of programming after watching the videos.

Chapter 6

Discussion

Our investigation found no differences in learning, cognitive load, social presence, interest, programming mindset, or self-efficacy between conditions of a monologue without the instructor visible, a monologue with the instructor visible, and a dialogue with both instructor and student visible.

6.1 Effects of speaker visual presence

The visual presence of the instructor in the monologue condition, and the instructor and student in the dialogue condition, were means of adding visual social cues to our 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 presentations, thereby facilitating better learning. On the other hand, a competing set of predictions was that the visual presence of the speaker(s) would be a seductive detail and would induce greater cognitive load, leading to impaired learning. We found that participants looked at the speakers in the visual presence conditions for about one fifth of the total trial time, indicating that the speakers in both monologue and dialogue conditions attracted participant focus. Despite this clear pattern of attention to the speakers, however, we found no differences in learning or feelings of social presence, interest, and cognitive load between either monologue or dialogue and the control. The same was true when we compared the dialogue to the monologue condition, suggesting that the presence of the student did not have an additional effect on participants' perceptions of the videos or learning outcomes.

The dwell time % that our participants spent fixating on the speakers in the videos — about 19% for monologue and 23% for dialogue — falls roughly in the middle of percentages measured the same way by other eye tracking studies: a higher proportion, 29.7%, was found by van Wermeskerken et al. (2018), followed by 24.8% in Colliot and Jamet (2018), and approximately 15% and 12% in two different visual presence conditions in van Wermeskerken and van Gog (2017). Our observed fixation % of 15% for monologue and 19% for dialogue were roughly comparable to van Gog et al. (2014), whose participants had approximately 20% of total fixations on the instructor. Only one of these studies, by Colliot and Jamet (2018), also measured social presence, interest, and cognitive load as we did, and found identical null effects of speaker visual presence in all measures. This gives us little precedent to which we may compare our results, but our findings do contribute to a pattern of visual presence not having an impact on social presence, interest, or cognitive load.

The observed effect of trial on attention to the speakers in our study can be compared to other eye tracking work that included two trials. In contrast to our finding that participants paid less attention to the speakers in the second trial, other participants have been found to spend a comparable (van Gog et al., 2014), or increased (van Wermeskerken and van Gog, 2017) amount of attention (note that only data from monologue studies are available). Our result suggests that the speakers were a seductive detail because participants focused less on them as the learning content difficulty increased. Meanwhile, the behaviour from participants in van Gog et al. (2014) and van Wermeskerken and van Gog (2017) suggests that the speaker was an important contributor to the learning content. This discrepancy is best explained by two significant ways that these studies differed from ours: one, they featured physical demonstrations as the learning content, and two, the same video was shown in both trials. The instructor featured in the videos was shown seated at a table, showing either how to complete a puzzle (van Gog et al., 2014) or construct a molecule (van Wermeskerken and van Gog, 2017) using small plastic pieces. In both cases, the instructor's face and gaze may have been important to attend to because it gave clear cues to when and where participants needed to focus on the demonstration area. In viewing the same demonstration a second time, participants in van Gog et al. (2014) and van Wermeskerken and van Gog (2017) may have still felt the need to attend to the instructor's face in order to pick up information they missed in the first viewing. In our study, instead of gaze, the mouse cursor of the instructor (and student, in dialogue) served as 'gestures' to direct participant attention since the speakers could not look directly at the learning content. Our participants may therefore have found the mouse cursors to be a more helpful cue in guiding their attention than the speaker(s), and thus felt a decreasing need to rely on the speaker(s) over time. Additionally, instead of showing the same content in both trials, our second trial consisted of content that built on the topics covered in the first. If participants perceived the speaker(s') presence as non-essential to learning, they may have chosen to dedicate less attention to the speaker(s) in the second trial, especially as topic difficulty increased.

The null conditional effect on cognitive load, social presence, and interest measures were mirrored by null conditional learning effects. In light of these results, it is somewhat striking that monologue and dialogue participants spent significantly less attention on the learning content than control participants but did not perform worse in terms of pre- to post-test gains. van Wermeskerken et al. (2018) 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 (2017) 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.

To our knowledge, only three eye tracking studies have observed higher learning outcomes in instructor-present conditions (Colliot and Jamet, 2018; van Gog et al., 2014; Wang and Antonenko, 2017). While promising, it is important to consider that the learning benefits reported in those works were limited to either certain types of information retention (Colliot and Jamet, 2018), repeated viewing of the learning content (van Gog et al., 2014), or an easy topic (Wang and Antonenko, 2017). In the broader context of visual presence studies, including those without eye tracking, the numerous null effect findings (Berner and Adams, 2004; Homer et al., 2008; Lyons et al., 2012; Wilson et al., 2018) make it difficult to conclusively argue that speaker visual presence is widely beneficial to learning. The modest sample sizes across the board (usually about 25 participants per condition for each of the aforementioned studies) is one limitation thus far for visual presence eye tracking studies; as such, revisiting our research setup with a larger sample size is one opportunity for further investigation.

Another noteworthy finding in comparing the effects of speaker visual presence was that participants in all three conditions spent similar amounts of time on the review screen. We had predicted that the potentially distracting nature of the speaker(s) would lead participants in the monologue and dialogue conditions to require more review time as compensation for attention not paid to the learning content. While we did find that participants in those conditions spent less time on the learning content, they did not then spend more time on the review screen, and ultimately, did not learn less. This echoes the finding of Colliot and Jamet (2018), who attribute this to participants "[knowing] how to regulate their behaviour" while watching the videos, such that they knew how to reduce disturbances to their learning and did not need to compensate for lost attention to learning content (p. 1429). If this is true, our study serves as an example that this self-regulation extends to people watching dialogue videos. But the significant effect of trial — where participants in all conditions spent less time on the second review screen than the first — warrants further discussion. One explanation for this reduced time on the second review screen is participant fatigue, where participants simply wanted to finish the experiment and leave. We did attempt to mitigate this effect by providing the five minute break in between the videos, which were themselves only about 10–15 minutes each. Another perspective is that the Trial 1 video was shorter than the Trial 2 video and spent less time explaining the material. The topics in the Trial 1 video, though less challenging than those in Trial 2, were reviewed in rather quick succession. Participants, especially if encountering programming for the first time, may have therefore felt the need to review the material more carefully once the video was over. In contrast, more of time and details were given to discussing the content in Trial 2, so by the time participants reached the review screen, they may have felt adequately versed in the content and less motivated to review it.

6.2 Effects of dialogue format

The second perspective from which to consider our results was that of the effects of dialogue format. We had predicted that the dialogue format would provide participants with a positive role model for their learning in the form of the student. Chi et al. (2017) had been a particular motivator of this hypothesis, as they postulated that better learning occurred from dialogue videos specifically because participants were more attentive to the student's contributions than the instructor's. Their findings were based on analysis of participants' utterances and not eye tracking behaviour, so we sought to add to this conjecture using eye tracking data. We found that participants paid less attention to the student than the instructor. The pattern of attention could be explained by the human tendency to choose others to learn from "who are likely to possess better-than-average information" (Henrich and Gil-White, 1999, p. 167), and to avoid those who are less skilled (Chudek et al., 2012). This phenomenon, known as prestige bias, would account for why the instructor garnered more attention from participants: she clearly possessed the most domain knowledge of the two dialogue speakers, and therefore would have made a more attractive model to learn from than the less experienced student.

We also found that dialogue condition participants did not learn more than monologue or control participants, did not report higher levels of social presence, interest, or cognitive load, and finally, did not attribute higher levels of social presence to the student than the instructor. Taken together, these findings provide little evidence to assert that the dialogue made any substantial impact on participants over the control or monologue.

Literature from dialogue studies provides some evidence for three key factors that may explain our null results. The first factor is the inclusion of deep questions —

questions that require either logical, causal, or goal-oriented reasoning to connect ideas (Bloom, 1956). For instance, the work by Driscoll et al. (2003), Craig et al. (2006), and Gholson et al. (2009) investigated the use of deep questions in scripted monologue and dialogue videos. They found that when deep questions were used in instructional videos, learning outcomes were enhanced regardless of delivery via monologue or dialogue. The same was found for the second factor, misconceptions: a dialogue in which the student expressed misconceptions and was corrected by the instructor produced better learning than a monologue without misconceptions, but a monologue where the instructor also covered misconceptions produced comparable learning gains in participants (Muller et al., 2007, 2008). Finally, the degree to which participants are interactive and collaborative while watching a dialogue was found to be yet another mediator of learning outcomes (Muldner et al., 2014).

Considering these factors together, we can begin to understand why Chi et al. (2017), r.f. Muldner et al. (2014), found significant impacts of dialogue on learning and attention but we did not. First, the unscripted videos used in Muldner et al. (2014) meant that questions and misconceptions were not controlled for. It is likely that the tutors, when delivering the monologue presentation, did not ask as many questions or address misconceptions as they did when interacting with a student in the dialogue presentation. As mentioned above, the use of both these features in instructional videos influences learning outcomes, so if the dialogue videos used in Muldner et al. (2014) contained more deep questions and misconceptions than the monologue, this could have skewed the results in favour of the dialogue. The scripting in our videos controlled for both these factors since we wanted to ensure that content was presented in the same way across conditions; as a consequence, the effect of deep questions and/or misconceptions would have been equalized between the monologue and dialogue conditions.

Second, participants in Muldner et al. (2014) were given workbooks and watched the videos in dyads, thus giving them the opportunity to talk about what they heard and observed with a peer while solving problems related to the learning material. According to Chi's ICAP (interactive, constructive, active, passive) framework, learning is largely dependent on how actively students engage with the learning materials and with each other (Chi, 2009; Chi et al., 2017). In the case of the Muldner et al. (2014) study, participants who watched a dialogue video were found to engage in more interactive and constructive (substantive) behaviour (i.e., ones that relate to the domain content), which correlated with higher learning gains than participants who watched a monologue video, or who engaged in fewer substantive behaviours. It could be that dialogue videos enhance learning only when opportunities are present that encourage interactive and constructive behaviour, but once those opportunities are absent — as in our study — so too are the learning advantages of dialogue. It is plausible that we may have seen similar results to Muldner et al. (2014) had we included more interactive or constructive elements in our study such as practice problems or collaborative activities. Thus, a potential limitation of the present study is the lack of interactivity with the learning content and opportunities for practice. This point certainly warrants further investigation since the findings of Chi et al. (2017) remain promising for guiding the design of effective learning from dialogue presentations.

6.3 Implications and future work

Overall, there is a substantial body of evidence suggesting that several elements contribute to learning and affective outcomes independently of visual presence (speaker(s) visible vs. not visible) or video format (monologue vs. dialogue). Authors have shown that visual preference (Homer et al., 2008), technological efficacy (Lyons et al., 2012), and topic difficulty (Wang and Antonenko, 2017) moderate how positively or negatively participants rate instructional videos with or without speaker visual presence. Similarly, deep questions, misconceptions, and interactivity have been shown to impact learning in dialogue research in spite of, and not because of, dialogue formatting. Because our study ensured that learning content and delivery was as similar as possible between conditions and that participants were relatively homogenous in their existing programming knowledge, we likely controlled for variables that would have produced conditional differences.

As far as providing explanations for our null learning and psychological results, the seductive detail and social presence frameworks were unable to provide clear answers. The fact that monologue and dialogue participants spent significantly less time attending to the learning content, yet did not report higher levels of cognitive load and did not learn less, shows little support for a seductive detail effect. Likewise, since participants showed no conditional differences on measures of social presence or interest and did not learn more in the conditions with speaker visual presence, we do not have support for a social presence effect. One perspective that could motivate future work is that elements of both seductive detail and social presence effects cancelled each other out. For example, any cognitive load induced by watching the speaker(s) or switching focus between the speaker(s) and the Python content may have been moderated by the visual social cues provided by the speaker(s). Futher investigation of this conjecture could compare the effects of visual seductive details (like a video of the instructor) with manipulations to the social cues, such as the use of a mechanical voice or the removal of gestures (thus reducing the amount of social cues provided). If learning outcomes were to be negatively impacted, this would provide evidence for social cues being a mediating factor to cognitive load and learning.

Another perspective that could explain the lack of conditional learning and psychological results is a lack of active engagement. As a reminder, instructional videos, when designed as essentially digital lectures, are susceptible to the same weaknesses as their traditional analogues. Namely, in order to effectively promote learning, they require elements that actively engage students. Without engagement, students are more likely to feel uninterested and in turn, learn less. The modest learning gains we observed in our participants may be symptomatic of this suppressed learning: from pre- to post-test and collapsed across condition, learning scores did improve significantly, and with a large effect size. However, post-test scores were not at all close to ceiling, with a total average of about 55%. The main manipulation in our videos, after learning content and delivery were controlled for, was the inclusion of the speaker(s). These alone likely did not contribute a great deal to the learning environment by way of inviting interaction from participants. In essence, whether or not participants saw the speakers, all the videos were pre-recorded so direct interaction between participants and speaker(s) was not possible. Participants were also not given anything to engage their learning as they watched the videos (like a worksheet to fill out), so they were only able to absorb the information passively.

Future work could build on these findings by comparing visual presence conditions alongside elements that enhance participant engagement, for example, through small-group discussion (van Blankenstein et al., 2011). Of particular interest would be any interactions that occur specifically between visual presence and higher interactivity, since prior work that has already shown promise of coupling dialogue with interactivity. Additionally, many of the variables we measured were informed by the seductive detail hypothesis and social agency theory, but these are just two frameworks of many. It is possible that other factors that we did not measure played a role in our results but went undetected, and so consideration of other measures is key to furthering the collective understanding of learning from instructional videos.

Several other avenues for exploration are open in terms of expanding this study's design: for instance, an experimental setup that more closely mirrors an at-home learning environment (e.g., allowing participants to pause, rewind, or fast-forward the video material; excluding eye tracking or using an eye-tracker that does not require participants to be situated in a headrest) may produce new insights. Replication in a classroom environment, or with a larger sample size, would also add robustness to the data. As demonstrated in this work, the use of eye tracking data has the potential to provide rich explanatory power to behaviour and effects that are otherwise only detectable through learning outcomes or self-reported measures. One application could be understanding how observers attended to the Python content being referenced — for example, when the instructor in the video referred to a certain line of code being displayed, did participants fixate on the correct line? Such investigation could yield interesting data about how novice learners attend to and understand code.

6.4 Conclusion

In conclusion, our study found no evidence that the visual presence of the speaker(s) in instructional videos was either beneficial or detrimental to learning: 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. Thus, our results gave little support for either seductive detail or social presence effects. However, given the possibility that the null effects were the result of seductive detail and social presence elements cancelling each other out, it may be premature to conclude that speaker visual presence has no effect at all on learning. Similarly, we were unable to show that learning from a dialogue video was more effective than learning from a monologue. Our manipulation of the dialogue ensured it was as similar as possible to the monologue format in terms of topics, questions, and misconceptions covered, so our results do not support the hypothesis that dialogues are superior per se.

Videos are likely to remain a popular choice in online learning. Instructional video producers must therefore be able to both recognise the limitations of the format and understand which factors produce optimal learning gains. This thesis has explored three video formats through measures of learning, visual attention, and five psychological measures. Our findings contribute to a more nuanced understanding of instructional video design and invite future examination of factors such as interactivity that can help the medium reach its full potential.

Bibliography

Adcock, A. B. and Van Eck, R. N. (2005). Reliability and factor structure of the attitude toward tutoring agent scale (ATTAS). *Journal of Interactive Learning Research*, 16(2):195–212.

Anderson, G. and Beal, C. R. (1995). Children's recognition of inconsistencies in science texts: Multiple measures of comprehension monitoring. *Applied Cognitive Psychology*, 9(3):261–272.

Atkinson, R. K., Mayer, R. E., and Merrill, M. M. (2005). Fostering social agency in multimedia learning: Examining the impact of an animated agent's voice. *Contemporary Educational Psychology*, 30(1):117 – 139.

Ayres, P. and Sweller, J. (2005). *The split attention principle in multimedia learning*, pages 135–146. Cambridge University Press, New York, NY.

Bandura, A. (1977). *Social Learning Theory*. Prentice-Hall, Englewood Cliffs, NJ.

Bandura, A. (1982). Self-efficacy mechanism in human agency. *American Psychologist*, 37(2):122–147.

Bandura, A., Ross, D., and Ross, S. A. (1961). Transmission of aggression through imitation of aggressive models. *Journal of Abnormal and Social Psychology*, 63(3):575–582.

Bates, A. W. (2015). *Teaching in a Digital Age*. BC Open Textbooks.

Berner, E. S. and Adams, B. (2004). Added value of video compared to audio lectures for distance learning. *International Journal of Medical Informatics*, 73(2):189 – 193.

Bligh, D. A. (1998). *What's the Use of Lectures?* Intellect, Exeter, UK.

Bloom, B. S. (1956). *Taxonomy of educational objectives: The classification of educational goals. Handbook I: Cognitive domain*. McKay, New York, NY.

Bloom, B. S. (1984). The 2 sigma problem: The search for methods of group instruction as effective as one-to-one tutoring. *Educational Researcher*, 13(6):4–16.

Chi, M. T. H. (2009). Active-constructive-interactive: A conceptual framework for differentiating learning activities. *Topics in Cognitive Science*, 1:73–105.

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

Chi, M. T. H., Roy, M., and Hausmann, R. G. M. (2008). Observing tutorial dialogues collaboratively: Insights about human tutoring effectiveness from vicarious learning. *Cognitive Science*, 32(2):301–341.

Chudek, M., Heller, S., Birch, S., and Henrich, J. (2012). Prestige-biased cultural learning: bystander's differential attention to potential models influences children's learning. *Evolution and Human Behavior*, 33(1):46–56.

Cohen, P. A., Kulik, J. A., and Kulik, C.-L. C. (1982). Educational outcomes of tutoring: A meta-analysis of findings. *American Educational Research Journal*, 19(2):237–248.

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

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

Cooper, K. M., Ding, L., Stephens, M. D., Chi, M. T. H., and Brownell, S. E. (2018). A course-embedded comparison of instructor-generated videos of either an instructor alone or an instructor and a student. *CBE Life Sciences Education*, 17(2):1–15.

Cox, R., McKendree, J., Tobin, R., Lee, J., and Mayes, T. (1999). Vicarious learning from dialogue and discourse: A controlled comparison. *Instructional Science*, 27(6):431–457.

Craig, S., Gholson, B., and Driscoll, D. M. (2002). Animated pedagogical agents in multimedia educational environments: Effects of agent properties, picture features, and redundancy. *Journal of Educational Psychology*, 94(2):428–434.

Craig, S. D., Chi, M. T. H., and VanLehn, K. (2009). Improving classroom learning by collaboratively observing human tutoring videos while problem solving. *Journal of educational psychology*, 101(4):779–789.

Craig, S. D., Sullins, J., Witherspoon, A., and Gholson, B. (2006). The deep-level-reasoning-question effect: The role of dialogue and deep-level-reasoning questions during vicarious learning. *Cognition and Instruction*, 24(4):565–591.

Driscoll, D. M., Craig, S. D., Gholson, B., Ventura, M., Hu, X., and Graesser, A. C.

(2003). Vicarious learning: Effects of overhearing dialog and monologue-like discourse in a virtual tutoring session. *Journal of Educational Computing Research*, 29(4):431–450.

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

Dweck, C. S. (1999). *Self-theories: Their role in motivation, personality, and development*. Essays in social psychology. Psychology Press, New York, NY, US.

Fox Tree, J. E. (1999). Listening in on monologues and dialogues. *Discourse Processes*, 27(1):35–53.

Gholson, B., Witherspoon, A., Morgan, B., Brittingham, J. K., Coles, R., Graesser, A. C., Sullins, J., and Craig, S. D. (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):487–493.

Gunawardena, C. N. (1995). Social presence theory and implications for interaction and collaborative learning in computer conferences. *International Journal of Educational Telecommunications*, 1(2):147–166.

Guo, P. J., Kim, J., and Rubin, R. (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, pages 41–50, New York, NY, USA. ACM.

Hansch, A., Newman, C., Hillers, L., Schildhauer, T., McConachie, K., and Schmidt,

P. (2015). Video and online learning: Critical reflections from the field. *HIIG Discussion Paper Series*, 2015(2):1–34.

Harp, S. F. and Mayer, R. E. (1998). How seductive details do their damage: A theory of cognitive interest in science learning. *Journal of educational psychology*, 90(3):414–34.

Hendy, H. M., Schorschinsky, N., and Wade, B. (2014). Measurement of math beliefs and their associations with math behaviors in college students. *Psychological assessment*, 26(4):1225–1234.

Henrich, J. and Gil-White, F. J. (1999). The evolution of prestige: freely conferred deference as a mechanism for enhancing the benefits of cultural transmission. *Evolution and Human Behavior*, 22(3):165–196.

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

Hyönä, J. (2010). The use of eye movements in the study of multimedia learning. *Learning and Instruction*, 20(2):172–176.

Jenkins, T. (2002). On the difficulty of learning to program. In *Proc. 3rd Ann. Conf. HEA Learning Teaching Support Netw.*, pages 53–58.

Just, M. A. and Carpenter, P. A. (1980). A theory of reading: From eye fixations to comprehension. *Psychological review*, 87(4):329–354.

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

Kizilcec, R. F., Papadopoulos, K., and Sritanyaratana, L. (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, pages 2095–2102, New York, NY, USA. ACM.

Koumi, J. (2006). *Designing Video and Multimedia for Open and Flexible Learning*. Routledge, Oxford, UK.

Krämer, N. C. and Bente, G. (2010). Personalizing e-learning: The social effects of pedagogical agents. *Educational Psychology Review*, 22(1):71–87.

Leppink, J., Paas, F., Van der Vleuten, C. P. M., Van Gog, T., and Van Merriënboer, J. J. G. (2013). Development of an instrument for measuring different types of cognitive load. *Behavior Research Methods*, 45(4):1058–1072.

Lester, J. C., Converse, S. A., Kahler, S. E., Barlow, S. T., Stone, B. A., and Bhogal, R. S. (1997). The persona effect: Affective impact of animated pedagogical agents. In *Proceedings of the ACM SIGCHI Conference on Human Factors in Computing Systems*, CHI '97, pages 359–366, New York, NY, USA. ACM.

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

Markman, E. M. (1979). Realizing that you don't understand: Elementary school children's awareness of inconsistencies. *Child Development*, 50(3):643–655.

Mayer, R. E. (2005a). *Cognitive theory of multimedia learning*, pages 31–48. Cambridge University Press, New York, NY.

Mayer, R. E. (2005b). *Principles for reducing extraneous processing in multimedia learning: Coherence, signaling, redundancy, spatial contiguity, and temporal contiguity principles*, pages 183–200. Cambridge University Press, New York, NY.

Mayer, R. E. (2005c). *Principles of multimedia learning based on social cues: Personalization, voice, and image principles*, pages 202–212. Cambridge University Press, New York, NY.

Mayer, R. E., Fennell, S., Farmer, L., and Campbell, J. (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):389–395.

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

Moreno, R., Mayer, R. E., Spires, H. A., and Lester, J. (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):177–213.

Muldner, K., Dybvig, K., Lam, R., and Chi, M. T. H. (2011). Learning by observing tutorial dialogue versus monologue collaboratively or alone. In *Proceedings of the 33rd Annual Conference of the Cognitive Science Society*, pages 1340–1346.

Muldner, K., Lam, R., and Chi, M. T. H. (2014). Comparing learning from observing and from human tutoring. *Journal of Educational Psychology*, 106(1):69–85.

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

Muller, D. A., Sharma, M. D., Eklund, J., and Reimann, P. (2007). Conceptual change through vicarious learning in an authentic physics setting. *Instructional Science*, 35(6):519–533.

Pressley, M., Ghatala, E. S., Woloshyn, V., and Pirie, J. (1990). Sometimes adults miss the main ideas and do not realize it: Confidence in responses to short-answer and multiple-choice comprehension questions. *Reading Research Quarterly*, 25(3):232–249.

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

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

Sweller, J. (2005). *Implications of Cognitive Load Theory for multimedia learning*, pages 19–30. Cambridge University Press, New York, NY.

Sweller, J., Ayres, P., and Kalyuga, S. (2011). *Intrinsic and extraneous cognitive load*, pages 57–69. Explorations in the Learning Sciences, Instructional Systems and Performance Technologies. Springer, New York, NY.

Sweller, J., van Merriënboer, J. J. G., and Paas, F. (1998). Cognitive architecture and instructional design. *Educational Psychology Review*, 10(3):251–296.

Sweller, J., van Merriënboer, J. J. G., and Paas, F. (2019). Cognitive architecture and instructional design: 20 years later. *Educational Psychology Review*, 31(2):261–292.

van Blankenstein, F. M., Dolmans, D. H. J. M., van der Vleuten, C. P. M., and

Schmidt, H. G. (2011). Which cognitive processes support learning during small-group discussion? The role of providing explanations and listening to others. *Instructional Science*, 39(2):189–204.

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

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

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

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

Wang, N., Johnson, W. L., Mayer, R. E., Rizzo, P., Shaw, E., and Collins, H. (2008). The politeness effect: Pedagogical agents and learning outcomes. *International Journal of Human-Computer Studies*, 66(2):98 – 112.

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

Wise, A., Chang, J., Duffy, T., and Valle, R. D. (2004). The effects of teacher social presence on student satisfaction, engagement, and learning. *Journal of Educational Computing Research*, 31(3):247–271.

Appendix A

Sample pre-test (administered as pen-and-paper test).

Pre questionnaire

Participant ID: ____________________ **Date: ______________**

Please answer the questions on the front and back. If you don't know, put IDK. Please show your work! For each of the following short programs, given the program (left) show what is printed (right).

```
result = 1
result = result + 10
print( result )
```

What does the program print?

```
name = input("enter your name: ")
print(" you entered", name )
```

User types **kat** - what does the program print?

```
char = "K"
letter = input("Enter a letter: ")
res = letter + char
print ( res )
```

User types **Z** - what does the program print?

```
result = 0
name = input("enter a city: ")
result = result + 1

name = input("enter another city: ")
result = result + 1

print( result )
print( name )
```

When answering the questions, the user types
Ottawa
TO
What does the program print? Show your work.

```
answer = input("Enter name: ")
if answer == "Bob":
    print(" hello ")
else:
    print(" hi ")
```

User types **Ann**. What does the program print?

```
val = 100
if val < 50:
    print("small")
else:
    print("medium")
```

What does the program print?

<pre>result = 1
while True:
    result = result + 1
    if result > 3:
        break
    else:
        print(result)
 	
print(" goodbye ")</pre>
	What does the program print? **Please <u>show your work</u>.**

<pre>answer = "king"
while True:
    user = input("Guess the card: ")
    if user == answer:
        print("yes!")
        break
    else:
        print("try again... ")</pre>
	The user types
	jack
	king
	What does the program print?

<pre>counter = 1
result = 0
while True:
    counter = counter + 1
    if counter > 2:
        break
    else:
        result = result + counter
print(counter, result)</pre>
	What does the program print? Please <u>show your work</u>.

Write a program that (1) continuously asks the user for 2 cities and (2) stops asking when the cities the user types are the same.

Appendix B

Self-efficacy questionnaire, delivered pre- and post-intervention. Responses were given on a scale from 1 (strongly agree) to 5 (strongly disagree).

1. I am confident that I can get a passing grade in a programming class.
2. I am confident that I can get an A in a programming class.
3. Programming seems easy to me, and I am confident I can get a good grade in a programming class.
4. Even if I do not understand a programming problem at first, I am confident I will get it eventually.
5. If I am taking a programming class and miss a lecture, I am confident that I can make up the work.
6. If I don't do well on a programming task, I know I can do better next time with more practice.
7. I am confident I can practice programming problems by myself until I understand them.

Appendix C

Programming mindset questionnaire, delivered pre- and post-intervention. Responses were given on a scale from 1 (strongly disagree) to 5 (strongly agree).

1. You can change even your basic programming ability level considerably.
2. No matter how much programming ability you have, you can always change it quite a bit.
3. You can learn new things, but you can't really change your basic programming abilities.
4. You can always substantially change your basic programming abilities.
5. To be honest, you can't really change how good at programming you are.
6. No matter who you are, you can significantly change your programming ability level.
7. You have a certain amount of programming ability, and you can't do much to change it.
8. Your programming ability is something about you that you can't change very much.

Appendix D

Situational interest questionnaire, delivered post-intervention. Responses were given on a scale from 1 (not true at all) to 5 (very true for me).

1. I enjoy working on programming.
2. I want to know more about programming.
3. I think programming is interesting.
4. I expect to master programming well.
5. I am fully focused on programming; I am not distracted by other things.
6. Presently, I feel bored.

Appendix E

Cognitive load questionnaire, delivered post-intervention. Responses were given on a scale from 1 (not at all the case) to 10 (completely the case). Items 1 and 2 measured intrinsic load, items 3, 4, and 5 measured extraneous load, and items 6, 7, 8, and 9 measured germane load.

1. The topics covered in the activity were very complex.
2. The activity covered formulas that I perceived as very complex.
3. The instructions and explanations during the activity were very unclear.
4. The instructions and explanations were, in terms of learning, very ineffective.
5. The instructions and explanations were full of unclear language.
6. The activity really enhanced my understanding of the topics covered.
7. The activity really enhanced my knowledge and understanding of programming.
8. The activity really enhanced my understanding of the formulas covered.
9. The activity really enhanced my understanding of concepts and definitions.

Appendix F

Social presence questionnaire, delivered post-intervention. Responses were given on a scale from 1 (not at all the case) to 6 (completely the case). Items 1 to 5 measured perceived speaker attitude toward the participant and items 6 to 9 measured participant interest/attention to the speaker. 'Instructor' was replaced with 'student' to additionally measure social presence of the student in the dialogue condition.

1. The instructor encouraged me to think for myself.
2. The instructor encouraged the development of my knowledge.
3. The instructor seemed friendly towards me.
4. The instructor seemed discouraging towards me.
5. The instructor seemed impatient with me.
6. The instructor made helpful comments.
7. The instructor increased my interest in the subject.
8. The teaching style of the instructor held my interest.
9. The instructor knows how to hold my attention when presenting material.

Appendix G

Excerpt of dialogue script.

PART 1

Slide 1
Instructor:
- This tutorial will introduce some basic foundations needed to write computer programs
- A program is a series of instructions that specifies how to perform some task, written in a programming language
- We will use the Python programming language
- Just to make this more concrete here is an example of a Python program

Student:
- Whoa, there's a lot going on here!

Instructor:
- It does look a little weird but that's because Python requires a specific syntax, or grammar, so that Python can understand what we are telling it to do.
- Don't worry if you don't understand this code now – it will be clear by the end of the lesson

Slide 2
Instructor:
- So as I said in the previous slide, this tutorial will go over the building blocks needed to write a program
- These are variables, including ways to print them and assign input from the user to them. We will also go over conditional statements and loops
- Before diving in, let's reflect on why we write programs. What do you think?

Student:
- I think it's so that we can make computers do things for us, like maybe if we want to find something quickly in a large collection of data?

Instructor:
- Exactly!

Slide 3
Instructor:
- We begin with variables
- You said programs operate on data. Variables let us label that data

Student:
- How do we create a variable in Python?

Instructor:
- There are two steps: we first write down the name of the variable and then we assign it a value
- Here is an example. We created a variable called Counter and assigned the value 0 to it

Student:
- Why is the variable called counter?

Instructor:
- That is the name we chose to give it – the name is up to us
- The syntax, or order of elements in our code, is important – Python wants us to have the name of the variable, followed by the assignment operator, followed by the value that is assigned to the variable, and all of that has to be on one line

Student:
- Is "assignment operator" just a fancy way of saying "equals sign"?

Instructor:
- Glad you asked – it's similar to an equals sign, but we should think of it as assignment, so 'counter' is getting *assigned* a value here

Appendix H

Consent form.

Carleton UNIVERSITY
Canada's Capital University

Consent Form

Title: Factors influencing learning from problem solving

Date of ethics clearance: Nov. 1, 2018, CUREB-B Clearance # 107272.

Ethics Clearance for the Collection of Data Expires: To be determined by the REB (as indicated on the clearance form)

I ____, choose to participate in a study that investigates factors influencing performance and learning from problem solving. **The researcher for this study is Kasia Muldner,** a professor in **the Institute of Cognitive Sciences,** as well as **Bridjet Lee**, a Carleton student being supervised by Muldner.

This study involves listening to a lecture on programming and solving some programming problems. No prior experience with programming is required. We will use an eye tracker to capture attention patterns. You will also be asked to fill in several questionnaires that will ask you questions (e.g., related to individual traits, attitudes towards programming, academic goals and personality). The study will take no more than 2 hours. As a token of appreciation, you will receive 2% extra credit for CGSC 1001. If you withdraw before finishing the experiment, the amount of compensation will be prorated by time.

You may feel some discomfort when you feel you don't know how to accomplish the experimental task. During the study session, you have the right to not answer any questions and/or end your participation in the study for any reason, by stating that you do not want to continue. If you withdraw from the study, all information you have provided will be immediately destroyed. Because we do not store any identifying information with the data, once you leave the session, withdrawal is not possible.

Research data will only be accessible by the researchers and no identifiable information will be recorded or stored. Thus, your name will not appear in any publications or other venues. Once the project is completed, all research data will be kept and potentially used for other research projects on this same topic.

If you would like a copy of the finished research project, please contact the researcher to request an electronic copy which will be provided to you.

The ethics protocol for this project was reviewed by the Carleton University Research Ethics Board, which provided clearance to carry out the research. If you have any ethical concerns with the study, please contact Dr. Bernadette Campbell, Chair, Chair, Carleton University Research Ethics Board-A (by phone at 613-520-2600, ext. 4085 or via email at ethics@carleton.ca).

Researcher contact information:
Name: Kasia Muldner
Department: Institute of Cognitive Sciences
Carleton University
Phone: 613-520-2600 x 2923
E-mail: kasiamuldner@cmail.carleton.ca

Name: Bridjet Lee
Department: Human Computer Interaction
Carleton University
E-mail: BridjetLee@cmail.carleton.ca

__________________________ _______________
Signature of participant Date

_________________________ _______________
Signature of researcher Date