Visual Augmentation of Deictic Gestures in MOOC Videos

Kshitij Sharma, Faculty of Business and Economics, University of Lausanne, kshitij.kshitij@unil.ch
Sarah D'Angelo, Northwestern University, sdangelo@u.northwestern.edu
Darren Gergle, Northwestern University, dgergle@northwestern.edu
Pierre Dillenbourg, École Polytechnique Fédérale de Lausanne, pierre.dillenbourg@epfl.ch

Abstract: We present an eye-tracking study to compare different modalities for visual augmentations of the teacher's explicit deictic gestures on a video lecture. We compared three visualizations: 1) hand gestures with a pointer, 2) gaze overlay, and 3) no-augmentation baseline. We investigate the teacher-student pair in a video-based learning context as an abstraction of an expert-novice pair where the goal is to attain a high level of shared understanding. The key phase of having a shared understanding is to have a common ground between the pair. Previous studies showed that explicit deixis plays a major role in initiating and maintaining common ground. This led us to hypothesize that augmenting videos with teacher's deictic gestures might help students perform better. We found that augmenting the video with teacher's gaze results in higher learning gain than no visualization. Moreover, gaze visualization also helped students in maintaining longer attention spans than hand gestures.

Keywords: eye-tracking, MOOCs, learning analytics, online deixis

Introduction

Lecture material can be displayed in a variety of formats to engage students and maintain their attention. Common augmentations include overlaying a pointer controlled by the teacher on the slides (Jermann, 2014) or including a video of the teachers face in the corner (Kizilcec et al., 2014). Now that eye tracking is becoming more accessible for educational applications, displaying teacher's gaze is another way to provide visual aids for students (Sharma et al., 2015).

Student's attention has been found to be correlated with performance in many studies related to visual tasks (Yantis and Jonides, 1984; Prinzmetal et. al., 1986; Juola et. al., 1991). In a visual comparison of two line segments Prinzmetal et. al. (1986) and Juola et. al., (1991) showed that the more attentive participants were more often correct in selecting the longer line segment. Yantis and Jonides (1984) found similar results in visual perception tasks. In a classroom, attention is: *"listening, sitting and working on assigned tasks" -Homes et. al. (2006).* In the context of academic performance previous research has shown strong association between students' attention and academic performance (Finn, 1989). In the context of video-based learning, one could use eye-tracking data to measure the amount of time the learner is paying attention to the elements that the teacher is referring to, verbally or through deictic.

In this study, we are particularly interested in maintaining attention in complex visual environments in which deictic references become important. We designed a video lecture on cloud identification to evaluate the presence of visual aids with highly visual and linguistically complex content. For this paper we will only consider the visual aspects of the content. In the case of a teacher-student dyad, following teacher's deictic references was correlated with the performance (Sharma et al., 2014). The main objective for any teacher-student dyad is to create a shared understanding of the content (a teacher-student dyad is a special case of the expert-novice dyad). Usually in dyadic interactions the basis of shared understanding is the common ground between the participants. Explicit deictic gestures play a key role in initiating and maintaining this common ground (Clark and Brennan, 1991).

In the present contribution, we use eye-tracking to capture student gaze patterns and compute their perceptual "with-me-ness" or the extent to which they follow along with the teacher's explicit deictic cues (Sharma et al, 2014); as well as their attention distribution across the lecture material. Additionally, we use posttest scores as an indicator of learning gain. Eye-tracking provides unprecedented access to the students' attention in video-based learning environments. Previous dual eye-tracking studies (Cherubini et al., 2008; Jermann and Nüssli, 2012; Richardson et al., 2007) have shown that eye-tracking could be used as an evaluation tool for the effectiveness of dyadic interaction. Cherubini et al (2008) found that the misunderstandings in a collaborative problem-solving task were correlated to the difference in participants' gaze patterns. Additionally, Richardson et al (2007) found that the cross-recurrence of a speaker-listener pair is correlated to the listener's comprehension. Furthermore, Jermann and Nüssli (2012) found that the cross-recurrence was correlated with the pair's collaboration quality. In this work we evaluate the presence of different deictic visualizations (pen pointer or gaze overlay) in a video lecture on learning gain and perceptual with-me-ness. We found that showing teacher's gaze to students improves learning gain compared to no visual aid. Additionally, the presence of gaze information increased perceptual with-me-ness compared to the pen pointer visualization.

Related work

The use of eye tracking in online education has provided researchers with insights about students' learning processes and outcomes. For example, Van Gog et al. (2005) used eye-tracking data to differentiate the expertise levels in the different phases of an electrical circuit-troubleshooting problem and concluded that experts focused more on the problematic area than the novices. In another experiment, where the participants had to learn a game, Alkan and Cagiltay (2007) found that the good learners focused more on the contraption areas (areas that appeared strange or unnecessarily complicated) of the game while they think about the possible solutions. Additionally, Mayer (2010) summarized the major results of research on eye tracking in online learning with graphics and concluded that there was a strong relation between fixation durations and learning outcomes; and visual signal guided students' visual attention. Understanding novice and expert gaze patterns in online education has informed a number of interventions to improve student learning. For example, Van Gog et al. (2009) found that displaying an expert's gaze during problem solving guided the novices to invest more mental effort than when there was no gaze displayed.

We know from previous eye-tracking research that speakers looked at the objects they refer to just before pointing and verbally naming the objects (Griffin and Bock, 2000). Listeners on the other hand, looked at the referred objects shortly after seeing the speaker point and refer to the objects (Allopenna et al., 1998). D.C. Richardson and Kirkham (2007) showed that the listeners who were better at attending the references made by the speaker were also better at understanding the context of the conversation. One way to aid the listeners attending the reference in a better way could be to display where the speaker is looking at. This might help the listeners in a better disambiguation of the complex references (Gergle and Clark, 2011, Hanna and Brennan, 2007). In the case of complex stimulus displaying the gaze of speaker made the disambiguation of the references even easier (Prasov and Chai, 2008).

Gaze contingent experiments are at the proactive side of the eye-tracking technology. These experiments consist in displaying the gaze of collaborating partners to each other; or displaying the gaze of an expert to a novice in order to teach the novice (Chetwood et al., 2012). Another modality of gaze contingency is using gaze as a mode of communication. In a collaborative "Qs-in-Os" search Brennan et al. (2008) showed that the sharing gaze information between collaborating partners resulted in a strategy of division of labor as effective as if the partners were talking face to face. Displaying the gaze of speaker helped the listener in deciphering the references (Gergle and Clark, 2011, Hanna and Brennan, 2007). Moreover, gaze of speaker made it easier for the listener in deciphering the references in situations with high ambiguity (Prasov and Chai, 2008).

Current study

Research question

Previous work has shown that with-me-ness is correlated with the learning gains (Sharma et al, 2014). When students were provided feedback on their with-me-ness and it improved their learning gains (Sharma, 2015). Additionally, it has been shown that putting teacher's gaze online correlates with a specific video navigation pattern (fewer and less frequent pauses, fewer backward jumps) that signifies the low perceived difficulty by the students (Sharma et al, 2015). However, these two results had two different modalities for augmenting the video with teacher's explicit deictic references: displaying the pointer and displaying the gaze. This study compares those two modalities with a baseline of no visual augmentation. Specifically, we address the following research question: How does deixis visualization affect the learning gain and students' attention?

Participants

Forty-three university students participated in the study; the mean age was 19.18 (sd = 2.07). Six of the participants were females and thirty-seven were males, all students were bachelor status. Informed consent was obtained for all participants and they received an equivalent of 20$ compensation for their participation.

Procedure

Participants were informed that they would be listening to a lecture on cloud identification and that their gaze would be recorded using a remote eye tracking system (SMI RED 250). Participants were asked to use a chin rest to keep their head stable and wear headphones. Before the experiment participants eye gaze was calibrated using a 5-point calibration. The start of the study began with ten-question cloud identification pretest, which was followed by the cloud identification lecture and concluded with ten-question cloud identification posttest. Subjects were informed that they should complete the tests and listen to the lecture at their own pace and they were able to pause the lecture and go back and forward in time as much as they wanted.

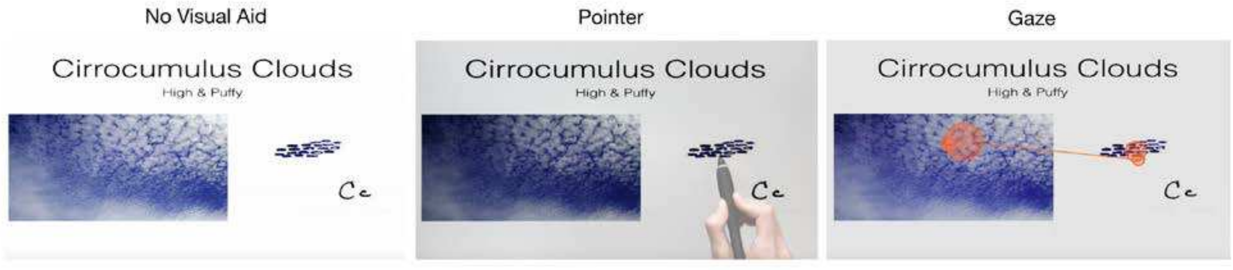
Task

The cloud identification pre and posttests consisted of ten multiple-choice questions. The ten cloud types covered in the lecture appeared on each test once, five of the cloud types were graphically represented and five were represented with photographs. The graphical representations and photograph representations were swapped for the pre and post test so each cloud type was represented both graphically and pictorially but no stimuli were repeated. Participants were asked to select the correct cloud type from four choices, they were instructed not to guess on the answers and if they did not know the correct answer to skip the question. The cloud lecture was 11 minutes and 37 seconds and consisted of seventeen slides. Each of the ten cloud types had an individual slide that contained the cloud name, two descriptors based on altitude and feature, and two representations a photograph and graphical depiction. The average time for a cloud content slide was 43 seconds (sd = 6.47 seconds). The lecture started and concluded with summary slides containing the ten types of clouds. The lecture content explained how to identify ten cloud types based on their distinguishing characteristics such as altitude in which they occur (i.e. high, medium, low) and describing features (i.e. puffy or layered). For example, the altocumulus cloud was described as a mid altitude cloud composed of puffy grey and white patches that are most likely to be seen with other clouds.

Measures

The participant's gaze was recorded for the duration of the study including the pretest, lecture, and posttest. All responses to the tests and interactions with the video (i.e. pauses) were recorded. Additionally, the gaze patterns for the teacher were recorded for the duration of the lecture, while the content was being recorded.

Independent variable


No Visual Aid

Pointer

Gaze

"…they look like a bunch of little grains arranged together…typically a group of very small elements"

Figure 1. Visual aid conditions.

Deixis visualization

We manipulated the availability of deixis visualization as a between subjects variable. There are three conditions of deixis visualizations: a pen pointer representation, gaze representation, or no additional visual aid (Figure 1). The lecture contained 242.6 seconds of pointer information; we replaced the same exact video segments with the gaze representation. The control condition does not contain any visual augmentation. Slide and lecture content were identical for all three conditions.

To determine the appropriate amount of gaze information to display, we conducted a small pilot study to evaluate how long the gaze trails should be visible on screen. Five participants viewed the lecture with both a 5 second gaze trail and a 2 second gaze trail alternating every 60 seconds of deixis visualization matching the pointer condition. Participants were asked if they noticed a difference in the gaze representation and if they did to state their preference for which gaze representation was most appropriate. The majority of the participants preferred the 2 second gaze trail, stating that the 5 second trail was disruptive and occluding too much content, therefore we used 2 second trail in the gaze condition.

Dependent variables

Learning gain

Learning gain was evaluated by the post test scores. Participants were not familiar with the content before the study and we observed a floor effect in the pre-test scores (median = 0, mean = 0.83/10, 27 out of 36 participants scored 0 or 1 in the pretest) therefore we do not consider pre-test scores in our analysis of learning gain. Since identifying clouds by name is a complex task, we developed a scoring rubric in order not to be restricted by the requirement of memorizing the cloud names. Students received one point for marking the correct name of the cloud. A half point was given to answers that had one of the correct characteristics of the cloud. For example if the correct answer is altocumulus students receive a half point for answers containing the prefix "alto" (indicating a mid-altitude cloud) or answers containing "cumulus" (indicating a puffy cloud shape). Zero points were given to incorrect answers that did not share correct characteristics.

Table 1: Example scoring grid for altocumulus cloud

	Cumulus or "Puffy"	**Stratus or "Layered"**	**"Exceptions"**	
High Altitude	Cirrocumulus (.5)	Cirrostratus (0)	Cirrus (0)	Cumulonimbus (0)
Middle Altitude	Altocumulus (1)	Altostratus (.5)	Nimbostratus (.5)	
Low Altitude	Cumulus (.5)	Stratus (0)	Stratocumulus (0)	

With-me-ness

With-me-ness is the extent to which the students succeed in following the teacher's dialogues and deictic gestures on the screen. In eye-tracking terms, with-me-ness captures: "how much time a student spent looking at the part of the display that the teacher is talking about?" With-me-ness is defined at two levels: perceptual and conceptual. There are two ways a teacher may refer to an object: with deictic gestures, generally accompanied by words ("here", "this variable") or only by verbal references ("the counter", "the sum"). Perceptual with-me-ness measured if the students looked at the items referred to by the teacher through deictic acts. Conceptual with-me-ness was defined using the discourse of the teacher: did students look at the object that the teacher was verbally referring to, i.e., that the teacher was referring to a set of objects that were logically or semantically related to the concept he was teaching. In the present study, since we are only interested in the effect of the deixis visualization on the students' gaze patterns, we will use the perceptual level of with-me-ness only. The perceptual "with-me-ness" has 3 main components: entry time, first fixation duration and the number of revisits. (a) Entry time was the temporal lag between the times a referring pointer appeared on the screen and stops at the referred site (x,y) and the time student first looked at (x,y). (b) First fixation duration was how long the student gaze stopped at the referred site for the first time. (c) Revisits were the number of times the student's gaze came back to the referred site.

Results

Time on video lecture: We observe no effect of time spent on the video-lecture on learning gain (r(43) = -0.03, p =.84). Moreover, we do not observe a difference in the time spent on the video between experimental conditions (F[2, 37] = 2.67, p = .10)

Learning gain: The learning gain in the gaze condition is significantly higher than the learning gain in no visual aid condition. However, there is no difference in learning gain across the pen pointer and the no visual aid condition. Table 2 shows the pairwise ANOVA results.

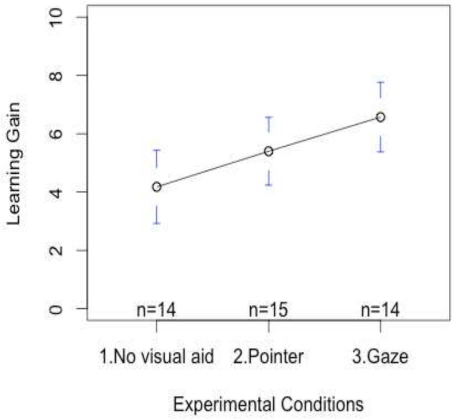


Figure 2. Learning gain.

Table 2: Pairwise ANOVA for learning gain

Condition pair	ANOVA effect size	p-value
Pointer vs. No visual aid	1.22	.28
Gaze vs. No visual aid	2.39	.01
Gaze vs. Pointer	1.17	.31

Perceptual With-me-ness: A one-way ANOVA, without the assumption for equal variances, shows that the perceptual with-me-ness is significantly higher in the gaze condition than that in the per pointer condition (gaze mean = 0.25, sd = 0.24, pointer mean = 0.09, sd = 0.09, no visual aid mean = 0.09, sd = 0.1, Figure 3). Table 3 shows the pairwise ANOVA results. In all the cases, i.e., pointer, gaze and no visual aid condition, we logged the exact point of the reference and then defined a circular area of 50 pixel diameter to compute the perceptual wit-me-ness.

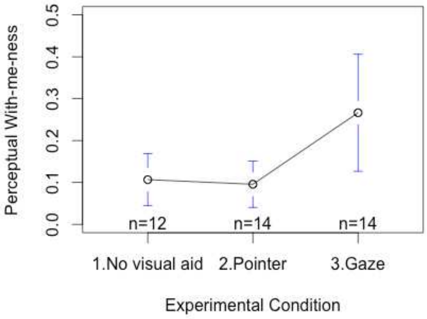


Figure 3. Perceptual With-me-ness.

Table 3: Pairwise ANOVA for perceptual with-me-ness

Condition pair	ANOVA effect size	p-value
Pointer vs. No visual aid	0.01	.98
Gaze vs. No visual aid	0.15	.05
Gaze vs. Pointer	0.17	.02

Gaze on video slides: We observe no difference in amount of time spent on the slides explaining individual cloud types, however, we do see a difference in the time spent looking at relevant content in the summary slides based on visualization condition. Participants who were shown the teacher's gaze overlay spend significantly more time looking at the specific cloud types in the summary slide compared to the no visualization condition and the pen pointer condition. There is no significant difference between the pen pointer condition and the no visualization condition. Table 4 shows the pairwise ANOVA results.

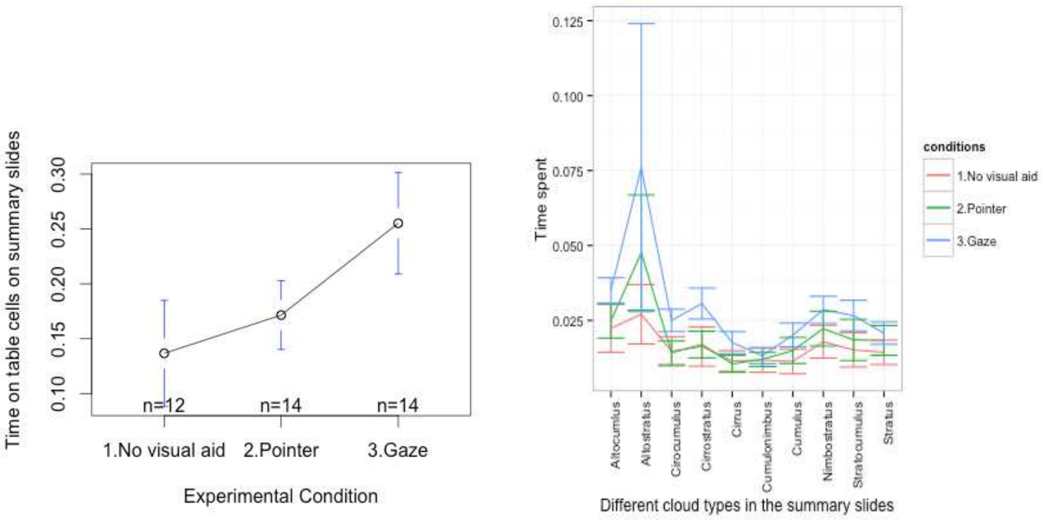


Figure 4. Time spent on summary slide and each cloud type.

Table 4: Pairwise ANOVA for time spent on summary slides

Condition pair	ANOVA effect size	p-value
Pointer vs. No visual aid	0.46	.50
Gaze vs. No visual aid	7.09	.01
Gaze vs. Pointer	6.02	.02

Discussion

The results of this study suggest that showing the teacher's gaze to students when making explicit references to information on the slides can be useful for students. As a visual aid, gaze highlights important areas on the slides that the teacher is explaining to students. Additionally, gaze provides more information than a pen pointer based representation, which may have contributed to the positive effects of sharing gaze information. Although we controlled for the time both visual aids were displayed on screen, the two second gaze trail allowed for multiple points of reference to be displayed at once while the pen was limited to a single point. For example, in one frame of the gaze condition the teacher can visually compare multiple areas on the slides. Additionally, gaze captures potentially unintentional signals that the teacher uses such as looking back at the name of cloud that may help students connect different knowledge points. This could have been less likely in the pen pointer condition, as the teacher intentionally controls the pointing behavior.

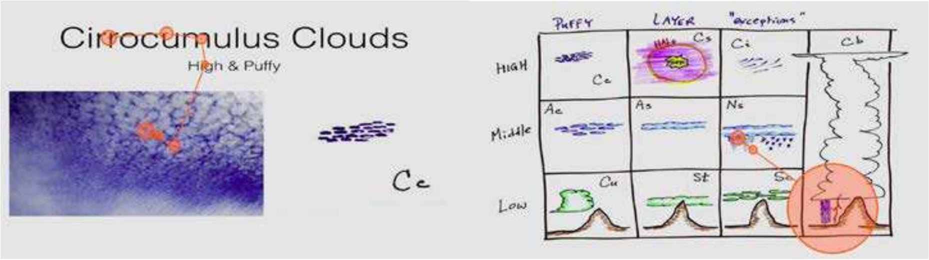


Figure 5. Examples of gaze use cases.

Since we do not see an effect of time spent on video, the increase in learning gain is most likely to be a result of where students were paying more attention in the lecture. We see that students spend more time looking at the cloud representations in the summary slides. This suggests that they may have spent more time comparing the distinguishing characteristics for cloud types, which could have contributed to improved performance on the posttest. The gaze representation may have been particularly useful, for the summary slides, because it is a more targeted representation and contains more temporal information about teacher's references. Whereas, the pen pointer enters from the bottom of the screen, which may distract attention to irrelevant areas of the slide. Another plausible explanation for the higher learning gain in the gaze condition could be higher levels of perceptual with-me-ness. Since we observe higher perceptual with-me-ness in the gaze condition; it suggests that students were following the teacher's gaze, which helped maintain attention to important part of the content. We see in Figure 3 that the overall level of perceptual with-me-ness is low (the students follow the teacher's references about 50% of the time at most). This shows that the students are not mechanically following the visual deictic, but using it as a support for following the teacher. These results are coherent with the results found by Sharma (2015).

Conclusions

Sharing teacher's gaze with students has a lot of potential for augmenting the videos with external information in online education. In environments like MOOCs gaze can be a practical addition to lecture content since the recording of content is not real time and eye tracking technology is becoming more accessible. Our results indicate that sharing gaze information improves learning gain compared to no visual aid and maintains students' attention for longer periods of time compared to the pen pointer condition. Therefore it could be a useful addition to visually rich and complex lecture content. Future analysis will investigate the relationship between linguistic complexity and student gaze given the visual aid representation. We will also investigate the amount of attention shift in the cloud specific slides and its effect on students' gaze on posttest items.

References

Alkan, S., & Cagiltay, K. (2007). Studying computer game learning experience through eye tracking. *British Journal of Educational Technology*, 38(3), 538-542.

Allopenna, P. D., Magnuson, J. S., & Tanenhaus, M. K. (1998). Tracking the time course of spoken word recognition using eye movements: Evidence for continuous mapping models. *Journal of memory and language*, 38(4), 419-439.

Brennan, S. E., Chen, X., Dickinson, C. A., Neider, M. B., & Zelinsky, G. J. (2008). Coordinating cognition: The costs and benefits of shared gaze during collaborative search. *Cognition*, 106(3), 1465-1477.

Cherubini, M., Nüssli, M. A., & Dillenbourg, P. (2008, March). Deixis and gaze in collaborative work at a distance (over a shared map): a computational model to detect misunderstandings. In Proceedings of the 2008 symposium on Eye tracking research & applications (pp. 173-180). ACM.

Chetwood, A. S., Kwok, K. W., Sun, L. W., Mylonas, G. P., Clark, J., Darzi, A., & Yang, G. Z. (2012). Collaborative eye tracking: a potential training tool in laparoscopic surgery. *Surgical endoscopy*, 26(7), 2003-2009.

Clark, H. H., & Brennan, S. E. (1991). Grounding in communication. Perspectives on socially shared cognition, 13(1991), 127-149.

Finn, J. D. (1989). Withdrawing from school. Review of educational research,59(2).

Gergle, D., & Clark, A. T. (2011, March). See what i'm saying?: using Dyadic Mobile Eye tracking to study collaborative reference. *In Proceedings of the ACM 2011 conference on Computer supported cooperative work* (pp. 435-444). ACM.

Griffin, Z. M., & Bock, K. (2000). What the eyes say about speaking. *Psychological science*, 11(4), 274-279.

Hanna, J. E., & Brennan, S. E. (2007). Speakers' eye gaze disambiguates referring expressions early during face-to-face conversation. *Journal of Memory and Language*, 57(4), 596-615.

Holmes, R. M., Pellegrini, A. D., & Schmidt, S. L. (2006). The effects of different recess timing regimens on preschoolers' classroom attention. Early Child Development and Care, 176(7).

Jermann, P., & Nüssli, M. A. (2012, February). Effects of sharing text selections on gaze cross-recurrence and interaction quality in a pair programming task. In Proceedings of the ACM 2012 conference on Computer Supported Cooperative Work (pp. 1125-1134). ACM.

Jermann, P. (2014). The EPFL MOOC Factory. Experience track contribution in EMOOCs 2014, Lausanne.

Juola, J. F., Bouwhuis, D. G., Cooper, E. E., & Warner, C. B. (1991). Control of attention around the fovea. Journal of Experimental Psychology: Human Perception and Performance, 17(1).

Kizilcec, R. F., Papadopoulos, K., & Sritanyaratana, L. (2014, April). 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* (pp. 2095-2102). ACM.

Mayer, R. E. (2010). Unique contributions of eye-tracking research to the study of learning with graphics. *Learning and instruction,* 20(2), 167-171.

Prasov, Z., & Chai, J. Y. (2008, January). What's in a gaze?: the role of eye-gaze in reference resolution in multimodal conversational interfaces. I*n Proceedings of the 13th international conference on Intelligent user interfaces* (pp. 20-29). ACM.

Prinzmetal, W., Presti, D. E., & Posner, M. I. (1986). Does attention affect visual feature integration? Journal of Experimental Psychology: Human Perception and Performance, 12(3), 361.

Richardson, D. C., Dale, R., & Kirkham, N. Z. (2007). The art of conversation is coordination common ground and the coupling of eye movements during dialogue. *Psychological science*, 18(5), 407-413.

Sharma, K., Jermann, P., & Dillenbourg, P. (2015). Displaying Teacher's Gaze in a MOOC: Effects on Students' Video Navigation Patterns. *In Proceedings of the 10th European Conference on Technology Enhanced Learning*.

Sharma, K. (2015). Gaze Analysis methods for Learning Analytics.

Sharma, K., Jermann, P., & Dillenbourg, P. (2014). "With-me-ness": A gaze-measure for students' attention in MOOCs. *In International conference of the learning sciences*.

Van Gog, T., Paas, F., van Merriënboer, J. J., & Witte, P. (2005). Uncovering the problem-solving process: cued retrospective reporting versus concurrent and retrospective reporting. *Journal of Experimental Psychology: Applied*, 11(4), 237.

Van Gog, T., Jarodzka, H., Scheiter, K., Gerjets, P., & Paas, F. (2009). Attention guidance during example study via the model's eye movements. *Computers in Human Behavior*, 25(3), 785-791.

Yantis, S., & Jonides, J. (1990). Abrupt visual onsets and selective attention: voluntary versus automatic allocation. Journal of Experimental Psychology: Human perception and performance, 16(1).