2019 IEEE 19th International Conference on Advanced Learning Technologies (ICALT)

Stimuli-based Gaze Analytics to Enhance Motivation and Learning in MOOCs

Kshitij Sharma[1], Pierre Dillenbourg[2], Michail Giannakos[1]
1. Norwegian University of Science and Technology, Trondheim, Norway
2. École Polytechnique Fédérale de Lausanne, Switzerland
Email: kshitij.sharma@ntnu.no; pierre.dillenbourg@epfl.ch; michailg@ntnu.no

Abstract—**The interaction with the various learners in a Massive Open Online Course (MOOC) is often complex. Contemporary MOOC learning analytics relate with click-streams, keystrokes and other user-input variables. Such variables however, do not always capture learners' learning and behavior (e.g., passive video watching). In this paper, we present a study with 40 students who watched a MOOC lecture while their eye-movements were being recorded. We then proposed a method to define stimuli-based gaze variables that can be used for any kind of stimulus. The proposed stimuli-based gaze variables indicate students' attention (i.e., with-me-ness), at the perceptual (following teacher's deictic acts) and conceptual levels (following teacher discourse). In our experiment, we identified a significant mediation effect of the two levels of with-me-ness on the relation between students' motivation and their learning performance. Such variables enable common measurements for the different kind of stimuli present in distinct MOOCs. Our long-term goal is to create student profiles based on their performance and learning strategy using stimuli-based gaze variables and to provide students gaze-aware feedback to improve overall learning process.**

Keywords: Eye-tracking, motivation, learning, MOOCs, video based learning, multimodal analytics, massive open online courses

I. Introduction

We present a study to investigate how well stimuli-based gaze analytics can be utilized to enhance motivation and learning in Massive Open Online Courses (MOOCs). Our work seeks to provide insights on how gaze variables can provide students with gaze-aware feedback and help us improve the design, interfaces and analytics used as well as provide a first step towards gaze-aware design of MOOCs to amplify learning.

The evidence for understanding and supporting users' learning is still very limited, considering the wide range of data produced when the learner interacts with a system (e.g., gaze [12]. Devices like eye-trackers have become readily available and have the capacity to provide researchers with unprecedented access to users' attention [19]. Thus, besides commonly used variables coming from users' click-streams, keywords and preferences, we can also use eye-tracking variables to accurately measure students' attention during their interaction with learning materials (e.g., MOOC lectures).

In this contribution, we address the general question of *how gaze-variables (related to students attention) can help students to watch MOOC videos more efficiently?*. We tackle this question from a teacher's perspective (how much student follows the teacher) and call it this gaze-based measure "with-me-ness". With-me-ness is defined in two levels: (1) perceptual (following teacher's deictic acts) and (2) conceptual (following teacher discourse). Specifically, in this contribution, we address ***how "with-me-ness" mediates the relationship between students' motivation and learning within a MOOC?***.

The rest of this paper is organised as follows. Section 2 outlines the relevant previous work. Section 3 illustrates the methodology used in the paper. Section 4 presents the results of the data analyses. Finally Section 5 discusses the results and concludes the paper.

II. Related work

a) ***Video based learning:***: The use of educational videos has been widely employed in the past years. Educational videos is a vital element in several online learning forms (in a MOOC, or how-to video tutorial), students spend enormous amount of time watching various forms of educational videos [15]. Educational videos have been studied extensively during the last decades, through the lenses of empirical studies and theories [4]. One of the most commonly acceptes theoretical angles it the one of the Cognitive Theory of Multimedia Learning (CTML) [10], CTML provides several insights on how video-based learning (and multimedia in general) can be used effectively.

Paivio [11] argued that information provided by both auditory and visual channels should increase recall and retention. Studies by Mayer [10] have shown that visual information helps to process and remember verbal information and vice versa. This argument was strengthened by cue-summation theory showing that learning performance in the combined audio and pictures was better than in the combined audio and text, if the numbers of available cues or stimuli are increased [16]. The major benefits of video as a facilitator of educational content include presentation of detailed information (with text and image), efficient engagement of students' attention, simulating discussions and providing concrete real life examples with visualizations [14].

During the last year, video-based learning practices are applied in a variety of ways, such as the flipped classroom, small private online courses (SPOCs), and xMOOCs. Today,

2161-377X/19/$31.00 ©2019 IEEE
DOI 10.1109/ICALT.2019.00052

advanced video repository systems have seen enormous growth (e.g. Khan Academy, PBS Teachers, Moma's Modern Teachers, Lynda) through social software tools and the possibilities to enhance videos on them [4].

Existing research on video-based learning involves many features of today's MOOCs lectures. Volery and Lord (2000) [21] identified 3 success factors in online education: usable and interactive technology design, instructors' enthusiasm and interest in the tool and students' exposure to the web. Tobagi (1995) [20] developed an online distant learning system to capture lectures real time, compress them, store them on an on-demand system and transmit the videos to internal server. The on-demand video system server eliminated distance limitations and provided time independent access to study material.

Video lectures have several affordances besides those relying to traditional fast-forward and rewind interactions. Innovative features, such as slide-video separation, social categorization and navigation, and advanced search, have also been used recently in video learning platforms [5]. All this amount of interactions can be converted via analytics into useful information that can be used to support learning [8]. As the number of learners and the diversity of collected data grows, our ability to capture richer and more authentic learning patterns grows as well, allowing us to create new affordances that amplify our learning capacity.

*b) **Eye-tracking and education:**:* Utilizing representative and accurate data allows us to better understand students and design meaningful experiences for them. Eye tracking has been employed to understand the learning processes and different levels of outcome in a multitude of learning scenarios. Prieto et al., [12] used eye-tracking data to explain the cognitive load that the teachers experience during different classes and various scenarios. These scenarios include different factors such as experience of the teacher, size of the class, presence of a new technology and presence of a teaching assistant. The results show that the eye-tracking data is an important source of information explaining different factors in teachers' orchestration load and experience.

Eye-tracking has also been used in online learning for both in individual [9] and collaborative [18] settings. Sharma et al., [19] focus on capturing the attention of the individual learners in a video-based instructional setting to find the underlying mechanisms for positive learning outcome; Sharma et al., [18] also focus on joint attention in remote collaborative learning scenarios to predict the learning outcome.

In general, eye-tracking allows us to generate rich information, but it can be challenging to identify what information is processed and retained based on human's gaze. The eye-mind hypothesis [7] proposes that there is a connection between people gaze and attention, assuming that people process the information that they visually attend to. In this contribution, we utilize eye-tracking to measure students attention and then address how students' attention (i.e., "with-me-ness") has the capacity to mediate the relationship between students' motivation and learning within a MOOC video.

III. METHODOLOGY

*a) **Participants and procedure**:* A total of 40 university students (12 females) from a major European university participated in the experiment. The only criterion for selecting the participant was that each participant took the introductory Java course in the previous semester. This is also a prerequisite for taking the Functional Programming in Scala course at the university campus. The participants watched two MOOC videos from the course "Functional Programming Principles in Scala" and answered programming questions after each video.

Upon their arrival in the experiment site the participants signed a consent form and answered the study processes questionnaire (SPQ) [2]. Then watched the two MOOC videos and answered the quiz based on what they were taught in the videos. During their interaction with the MOOC videos their gaze was recorded, using SMI RED 250 eye-trackers.

*b) **Measures**:* The measures used in our study were: students/teacher co-attention (i.e., with-me-ness) coming from eye-tracking, students motivation coming from SPQ and students learning (coming from the final test).

With-me-ness measures how much the student is paying attention to what the teacher is saying or pointing at [17], [19]. With-me-ness is defined at two levels of teacher-student interaction: perceptual and conceptual.

Perceptual with-me-ness measures if the student looks at the items referred to by the teacher through deictic acts (sometimes accompanied by words like, here, this variable or only by verbal references, like, the counter, the sum). Deictic references are implemented by using two cameras during MOOC recording, one that captures the teacher's face and one, above the writing surface, that captures the hand movements. In some MOOCs, the hand is not visible but teacher uses a digital pen whose traces on the display (underlining a word, circling an object, adding an arrow) act as a deictic gestures. The perceptual "With-me-ness" has 3 main components: entry time, first fixation duration and the number of revisits (Figure 1). Entry time (Figure 1 top-right) is the temporal lag between the times a referring pointer appears on the screen and stops at the referred site (x,y) and the first time the student's gaze stops at (x,y). First fixation duration (Figure 1 bottom-left) is how long the student gaze stops at the referred site for the first time. Revisits (Figure 1 bottom-right) are the number of times the student gaze comes back to the referred site. The measure of perceptual with-me-ness is an arithmetic combination of these components (FFD = First Fixation Duration; ET = Entry Time; NumRV = Number of revisits; RV = ReVisit duration):

$$\frac{FFD - ET + \sum_{i=1...NumRV} RV\_i}{Total\_duration\_of\_the\_deictic\_reference} \qquad (1)$$

Conceptual with-me-ness is defined by the discourse of the teacher (i.e., to what extend students look at the object that the teacher is verbally referring to) Figure 2 provides an example. Thus, conceptual with-me-ness measures how often a student looks at the objects verbally referred to by the teacher during the whole course of time (the complete video). In order to

Fig. 1. A typical example of following the teacher's deictic gestures in the video lecture.

have a consistent measure of conceptual "With-me-ness" we normalize the time a student looks at the overlapping content by slide duration.

c) Motivation: We used the motivation scales from the SPQ [2]. This is a 5-point Likert scale questionnaire containing 10 questions (5 for deep and 5 for surface motivation). Deep motivation is defined as having the intrinsic motivation towards learning, while the surface motivation is defined as fear of failing in the tests [2].In this study we used the mean motivation (mean of deep and surface) that has an average value of 2.10 (Std. Dev. = 1.20).

d) Learning: At the end of the videos the students took a test about the content they were taught in the two videos. The score form this test was considered to be the learning performance in this paper. After this point, we would refer to this as learning. The mean learning value was 6.9 out of 10 (Std. Dev. = 1.6). For the test, the instructor of the MOOC helped the authors to create the multiple choice quiz for the two videos. This quiz was similar to the one used in the MOOC running at Coursera platform.

e) ***Data Analysis****:* To identify how "with-me-ness" (measured by eye-tracking) mediates the relationship between students' motivation (measured by the questionnaire) and learning (measured by the post quiz) within a MOOC we employ mediation analysis proposed by Baron and Kenny [1]. In our analysis, we used motivation as the predictor, learning as the outcome and "with-me-ness" as the mediating variables. Figure 4 shows the schematic representation of the model.

To examine with-me-ness capacity to mediate the relationship between motivation and learning we followed Baron and Kenny [1] three steps process: a) the predictor (i.e., motivation) must significantly influence the mediator (i.e., with-me-ness); b) the predictor (i.e., motivation) must significantly influence the outcome (i.e., learning); c) both predictor and mediator are employed to predict the outcome: if both of them significantly affect the dependent variable, then this mediator partially mediates the impact of the predictor independent variable on the outcome; if the influence of mediator is significant but the influence of predictor is not, then mediator fully mediates the impact of predictor on outcome.

IV. Results

To examine the mediation effect of with-me-ness we tested the model shown in Figure 4 with both the perceptual and the conceptual variables of with-me-ness. As shown in Table I, the direct link between motivation and both variables of with-me-ness was significant and hence satisfied the first condition. The link between motivation and and learning was also significant and hence satisfied the second condition as well. Moreover, the direct relationship between motivation with learning was not significant when with-me-ness variables (perceptual and the conceptual) were added. In table I we present the results of the two mediation analyses (one for perceptual and one for conceptual with-me-ness).

We observe that learning can be significantly predicted by motivation and that perceptual with-me-ness can be predicted by motivation. Finally, there is a significant prediction of learning by motivation and perceptual with-me-ness, however the coefficient of motivation is not significant anymore. Thus we can conclude that the perceptual with-me-ness fully mediates the relation between motivation and learning. It is clear from Figure 3 that the students with high motivation have higher chances of getting a high score if they high perceptual with-me-ness than the students with lower motivation.

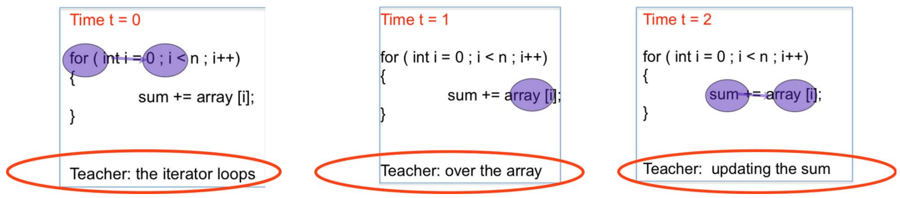

Fig. 2. A typical example of following the teacher's speech in the video lecture.

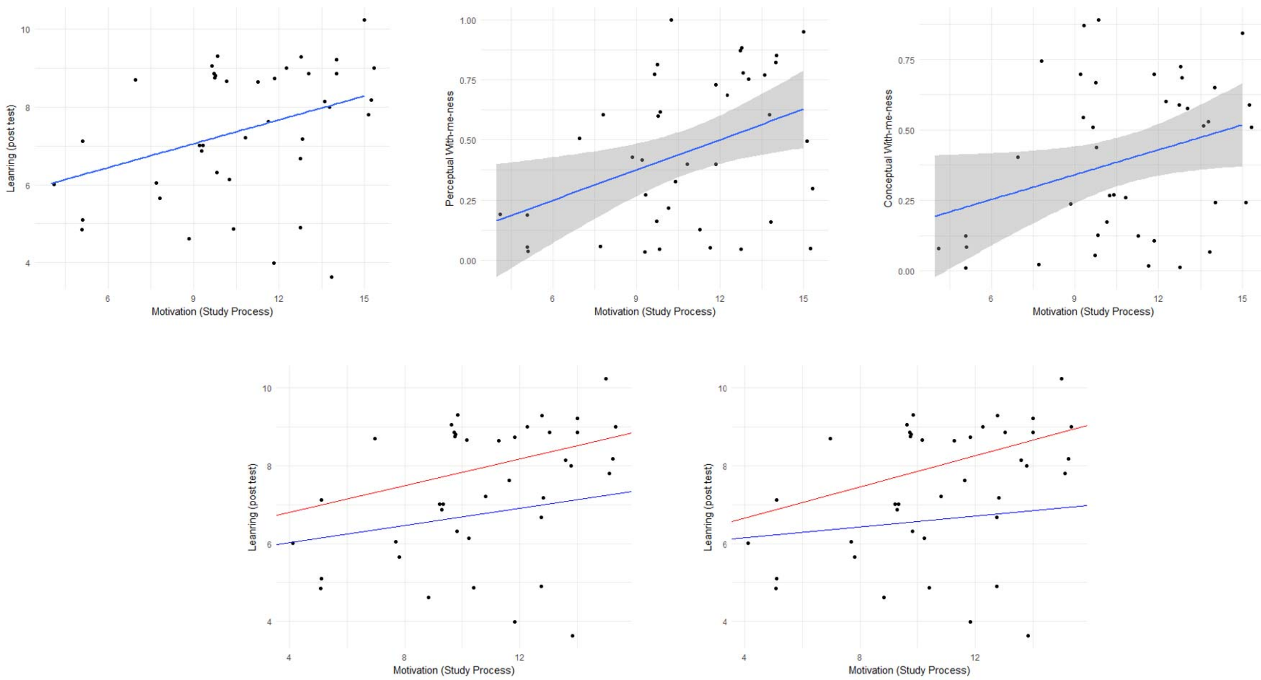

Fig. 3. Top left: learning predicted by motivation. Top-middle: perceptual with-me-ness predicted by motivation. Top-right:conceptual with-me-ness predicted by motivation. Bottom left: learning predicted by motivation (red = high perceptual with-me-ness, blue = low with-me-ness). Bottom right: learning predicted by motivation (red = high conceptual with-me-ness, blue = low conceptual with-me-ness).

TABLE I
MEDIATING EFFECT TESTS ($* \, p < .05$; $** \, p < .01$)

Predictor (Pr)	**Mediator (M)**	**Outcome (O)**	**Pr $\rightarrow$ M**	**Pr $\rightarrow$ O**	**Pr + M $\rightarrow$ O**		**Mediating effect**
					Pr	**M**	
Motivation	Perceptual with-me-ness	Learning	2.69**	2.40*	1.40	2.30*	Fully Mediated
Motivation	Conceptual with-me-ness	Learning	2.05*	2.40*	1.57	2.90 **	Fully Mediated

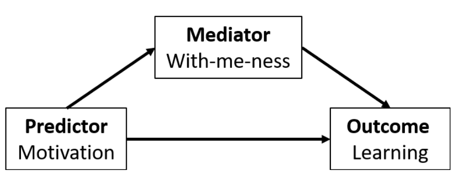

Fig. 4. Schematic representation of mediation effect and the example from the present contribution.

Next, we observe that that learning can be significantly predicted by motivation, and that conceptual with-me-ness can be predicted by motivation. Aslo, there is a significant prediction of learning by motivation and conceptual with-me-ness, however the coefficient of motivation is not significant anymore. Thus we can conclude that the conceptual with-me-ness fully mediates the relation between motivation and learning. It is clear from Figure 3 that the students with high motivation have higher chances of getting a high score if they high conceptual with-me-ness than the students with lower motivation.

V. DISCUSSION AND CONCLUSIONS

The reported study developed and empirically explored two models, where teacher/student co-attention (i.e., with-me-ness) were found to mediate the relationship of motivation and learning in MOOC videos. These two models demonstrated how the aspect of co-attention, not only influences learning, but also affects the effect of motivation in learning. Quantifying an often overlooked element (i.e., instructor's capacity to draw student's attention) in online courses.

In addition, we found that high-performers (those who scored high in the test) had more perceptual with-me-ness on the referred sites than the low-performers. This is in accordance with the literature, where Jermann and Nšsli [6], showed that better performing pairs had more recurrent gaze patterns during the moments of deictic references. We also found that the students who scored better in the test, were following the teacher, both in deictics and discourse, in an efficient manner than those who did not score well in the test. These results were not surprising, but could be utilised to inform the students about their attention levels during MOOC lectures in an automatic and objective manner. The results also contribute towards our long-term goal of defining the student profiles based on their performance and motivation using the gaze data. The attention points can serve the purpose of a delayed feedback to the students based on their attention span.

The conceptual with-me-ness can be explained as a gaze-measure for the efforts of the student to sustain common ground within the teacher-student dyad. Dillenbourg and Traum [3] and Richardson et al. [13] emphasised upon the importance of grounding gestures to sustain shared understanding in collaborative problem solving scenarios. A video is not a dialogue; the learner has to build common grounds, asymmetrically, with the teacher. The correlation we observed between conceptual with-me-ness and the test score (0.36, $p < 0.05$) seemed to support this hypothesis.

Another interesting finding of our study, is that the conceptual with-me-ness has more percentage mediation than the perceptual with-me-ness (39% for conceptual as compared to 33% for perceptual with-me-ness). This shows that eye-tracking can not only provide access to students' attention but also to the students' information processing mechanisms as well. Thus, students gaze is an important source of information that can be used to inform online learning.

To gain further insight into the design of MOOC videos and the affordances of the respective systems, we need to consider eye-gaze measurements (or can call them gaze analytics) that we found to not only strongly associated with learning, but also mediate the influence of other variables (i.e., motivation). Discussing these features from a technical standpoint can give rise to practical implications for the design of MOOC videos (e.g., designed in a way to draw students' attention [9]) and the respective video-based learning systems (e.g., offer an indication of students' attention based on the web-camera).

For future work, we are now beginning to collect eye-tracking data from different types instruction (e.g., pair problem solving) utilizing different stimulus (e.g., not controlled from the student like the video). In addition, we intend to investigate whether a plausible association exists between different students (e.g., novices). After identifying the role of with-me-ness and other gaze-analytics in different contexts, we will be able to propose how gaze-analytics can be integrated to various contemporary learning systems. For example, allowing us to enable student profiles based on their performance and learning strategy using gaze-analytics, and ultimately provide gaze-aware feedback to improve the overall learning process.

REFERENCES

[1] Reuben M Baron and David A Kenny. The moderator–mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. *Journal of personality and social psychology*, 51(6):1173, 1986.

[2] John Biggs, David Kember, and Doris YP Leung. The revised two-factor study process questionnaire: R-spq-2f. *British Journal of Educational Psychology*, 71(1):133–149, 2001.

[3] Pierre Dillenbourg and David Traum. Sharing solutions: Persistence and grounding in multimodal collaborative problem solving. *The Journal of the Learning Sciences*, 15(1):121–151, 2006.

[4] Michail N Giannakos. Exploring the video-based learning research: A review of the literature. *British Journal of Educational Technology*, 44(6):E191–E195, 2013.

[5] Michail N Giannakos, Konstantinos Chorianopoulos, and Nikos Chrisochoides. Making sense of video analytics: Lessons learned from clickstream interactions, attitudes, and learning outcome in a video-assisted course. *The International Review of Research in Open and Distributed Learning*, 16(1), 2015.

[6] Patrick Jermann and Marc-Antoine Nüssli. 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*, pages 1125–1134. ACM, 2012.

[7] Marcel A Just and Patricia A Carpenter. A theory of reading: From eye fixations to comprehension. *Psychological review*, 87(4):329, 1980.

[8] Juho Kim, Phu Tran Nguyen, Sarah Weir, Philip J Guo, Robert C Miller, and Krzysztof Z Gajos. Crowdsourcing step-by-step information extraction to enhance existing how-to videos. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 4017–4026. ACM, 2014.

[9] René F Kizilcec, Kathryn Papadopoulos, and Lalida Sritanyaratana. 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*, pages 2095–2102. ACM, 2014.

[10] Richard E Mayer and Roxana Moreno. Nine ways to reduce cognitive load in multimedia learning. *Educational psychologist*, 38(1):43–52, 2003.

[11] Allan Paivio. *Imagery and verbal processes*. Psychology Press, 2013.

[12] Luis P Prieto, Kshitij Sharma, Lukasz Kidzinski, and Pierre Dillenbourg. Orchestration load indicators and patterns: In-the-wild studies using mobile eye-tracking. *IEEE Transactions on Learning Technologies*, 2017.

[13] Daniel C Richardson, Rick Dale, and Natasha Z Kirkham. The art of conversation is coordination common ground and the coupling of eye movements during dialogue. *Psychological science*, 18(5):407–413, 2007.

[14] Daniel L Schwartz and Kevin Hartman. It is not television anymore: Designing digital video for learning and assessment. *Video research in the learning sciences*, pages 335–348, 2007.

[15] Daniel T Seaton, Yoav Bergner, Isaac Chuang, Piotr Mitros, and David E Pritchard. Who does what in a massive open online course? *Communications of the ACM*, 57(4):58–65, 2014.

[16] Werner Severin. Another look at cue summation. *AV Communication Review*, 15(3):233–245, 1967.

[17] Kshitij Sharma, Daniela Caballero, Himanshu Verma, Patrick Jermann, and Pierre Dillenbourg. Looking at versus looking through: A dual eye-tracking study in mooc context. In *Proceedings of 11th International Conference of Computer Supported Collaborative Learning, Gothenburg, Sweden, CSCL*, 2015.

[18] Kshitij Sharma, Daniela Caballero, Himanshu Verma, Patrick Jermann, and Pierre Dillenbourg. Shaping learners' attention in massive open online courses. *Revue internationale des technologies en pédagogie universitaire/International Journal of Technologies in Higher Education*, 12(1-2):52–61, 2015.

[19] Kshitij Sharma, Patrick Jermann, and Pierre Dillenbourg. "with-me-ness" : A gaze-measure for students? attention in moocs. In *International conference of the learning sciences*, 2014.

[20] Fouad A. Tobagi. Distance learning with digital video. *IEEE MultiMedia*, 2(1):90–93, 1995.

[21] Thierry Volery and Deborah Lord. Critical success factors in online education. *International journal of educational management*, 14(5):216–223, 2000.