Check for updates

Using the YouTube Video Style in a MOOC

(Re-)Testing the Effect of Visual Experience in a Field-Experiment

Hendrik Steinbeck, Theresa Zobel, Christoph Meinel
[firstname.lastname]@hpi.de
Hasso Plattner Institute
Potsdam, Brandenburg, Germany

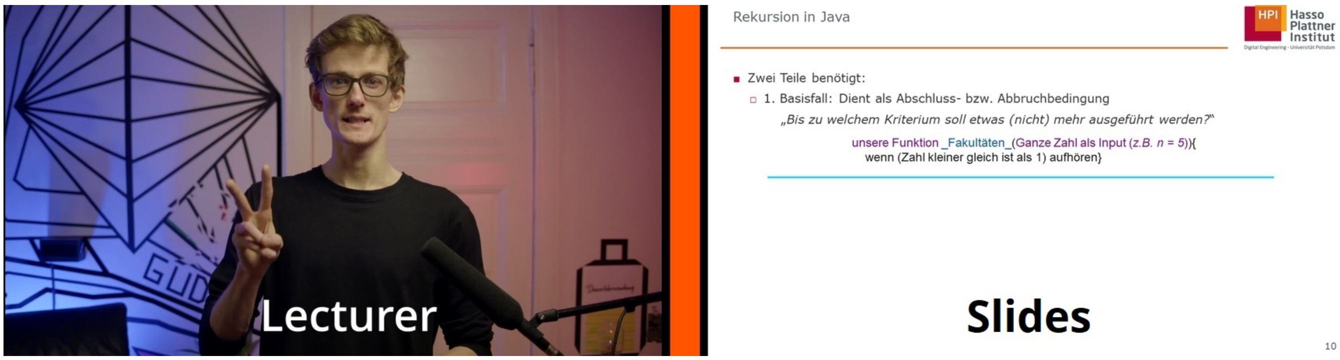

Figure 1: Experiment conditions at 1 minutes and 38 seconds: Informal explainer video (left) as seen on major video platforms, compared to traditional slide-based teaching (right).

ABSTRACT

Whether a lecturer presence helps or hinders in audio-visual learning material has been raised numerous times. While previous research found no substantial evidence in favor of a lecturer presence in controlled eye-tracking experiments, the given study analyzes results of a field-experiment in a German MOOC with 2,938 active participants taking a four-week course on data structures and algorithms. The research team produced specific content for this experiment with the goal to compare traditional slide-lectures with a modern explainer video style as seen on major video platforms. These two treatment groups are identical on the audio track and truly only differ in terms of the visual experience of the lecture on recursion. The variables are: 1) Perception of a learner defined by content, the speaker and his/her own learning and 2) Scores of a recall and transfer assessment. The first variable is conducted by a user survey (n=490), the skill assessment is measured by two posttests (quiz & programming task). The findings indicate a different perception of the speaker's focus, a significant better evaluation of the lecturer condition and higher scores on the recall posttest. No difference is seen in the perceived degree of professionalism, the self-reported level of attention and the scores of the transfer task. By testing and replicating previous findings in a real MOOC setting with adult learners, the given study contributes to the research of video-based learning in general, and to the sub-topic of effective teaching settings for computer science concepts in scaling environments.

This work is licensed under a Creative Commons Attribution International 4.0 License.

L@S '22, June 1–3, 2022, New York City, NY, USA.
© 2022 Copyright held by the owner/author(s).
ACM ISBN 978-1-4503-9158-0/22/06.
https://doi.org/10.1145/3491140.3528268

CCS CONCEPTS

• **Applied computing** → **Interactive learning environments**; **Distance learning**; • **Social and professional topics** → **Adult education**; • **Human-centered computing** → **Field studies**.

KEYWORDS

lecturer presence, video instruction, YouTube explainer video, production styles

ACM Reference Format:
Hendrik Steinbeck, Theresa Zobel, Christoph Meinel. 2022. Using the YouTube Video Style in a MOOC: (Re-)Testing the Effect of Visual Experience in a Field-Experiment. In *Proceedings of the Ninth ACM Conference on Learning @ Scale (L@S '22), June 1–3, 2022, New York City, NY, USA.* ACM, New York, NY, USA, 9 pages. https://doi.org/10.1145/3491140.3528268

1 INTRODUCTION

The popularity of the video as a media is unbroken and can be seen in various forms and styles: From shorter time frames for a video to reach 100 million views, to the growing watch time of live streaming content on various platforms (e.g. Twitch see [39]), to a higher number of daily active users in recent years [9]. The pandemic situation with its planned and unplanned behavioral changes in terms of how we work, play and teach can be seen as one driving factor behind the already existing trends of video usage and consumption. In the field of online education, rising numbers in the context of the pandemic can be witnessed as well (Key-Notes on L@S 2021, e.g. by Jeff Maggioncalda [23]). With this increase, new formats and lecture styles evolve, that need evaluation on a qualitative and quantitative basis. The given paper conducted a field-experiment design, comparing two different treatment conditions: The traditional slide-based teaching *against* a talking head format, that is seen in MOOCs and on general-purpose video platforms such as YouTube. By incorporating existing research designs and their findings, our study adds to the on-going question whether it's worth to include a visible lecturer presence or not. This kind of comparison seems worthwhile, as previous research did not find a significant difference regarding learning performance when it comes to the presence of an instructor. As the production of a video-based learning items needs resources – and in case of a visible lecturer – a face to show, the outcomes of the study is not only relevant to the community of scaling learning environments, but also to instructors and higher education professionals. The existing research foundation will be presented next, followed by the research questions and hypotheses as derivatives.

2 RELATED WORK

A first branch of the literature addresses the question of whether showing the lecturer has a positive or negative impact on student perception and learning. Multiple researchers conducted experiments using eye-tracking [4, 6, 20, 29, 38, 42]. The controlled laboratory environment can be seen as an advantage for measuring cause-effect relationships. Naturally, only a smaller number of participants can be tracked and analyzed in detail. The number of participants in the aforementioned studies ranges from 16 to 88, with the exception of Kokoç et al. [21] with 201. All of these studies focus on university or college students. A unique insight is provided by Kizilcec et al. [19], who analyze two large cohorts of participants in separate non-laboratory studies. They compare a strategic embedding with a constant lecturer presence, reporting a learners preference of seeing the lecturer. Two concepts are frequently cited in the literature: Social presence and the split-attention-effect. These two simultaneously form the most important poles for and against showing a lecturer.

2.1 Split-attention-effect

Following the theory of cognitive load, a learner can only comprehend a certain amount of information in a given time. The presence of a lecturer's video stream could add to the load. As a result, the pedagogical value of the socio-technical interaction via asynchronous video could affect learning [17, 24, 25]. Previous comparisons were often based on slides. The treatment conditions show a lecturer within the slides in a Picture-in-Picture (PiP) style [5, 19, 28]. It can be argued, that the split-attention-effect is higher in that comparison. When a learner has to switch between reading a slide and looking at the lecturer, the center of attention actually switches more frequently [6, 20]. A counter-argument is provided by Uchidiuno et al. and their large field-study, showing a relevance of text in videos from English language classes. [41].

2.2 Social presence

Social presence refers to a person's recognition in a medium [11]. Some argue that interaction, a human gaze or any other kind of para-social activity fosters learning [14, 44]. As Wang and Antonenko [43] found, easier learning items with a lecturers' condition are perceived better than more difficult ones. Homer et al. [16] argue that individual visual preferences influence the extent of perceived cognitive load. Regarding reliability, Ng and Przybyłek [28] have outlined that measuring social presence through surveys has its pitfalls, as it is prone to subjectivity. When only the face is visible, previous research has raised the question of the added value of showing the face compared to just listening [13]. A common answer are social cues and nonverbal communication that justify a talking-head format.

2.3 MOOC video styles

A second branch of the literature focuses on the video style produced for a MOOC. Those practices and taxonomies influences the production of treatment conditions: The results of Guo et al. [12] were taken into account when planning the videos. As a result, for the *Lecturer* treatment condition, we chose a short length (<6 min), a talking head format with a wide-angle to allow gestures and our own recording in an informal setting rather than an existing sequence of pre-recorded university instructional material recorded in a studio environment. Reutemann [32] analyzed four major MOOC platforms and found that a majority (74%) of courses use a talking head format, followed by slides (>33%) in several combined variants. In addition to PowerPoint variants, other dominant video styles include chalk lectures, screencasts, hands-on demonstrations and handwriting variants, as studied by Santos et al. [35]. All attempts to collect video style archetypes involve listing of non-exclusive elements; for example a talking head may use a voice-over, a drawing board view can have text overlays. For the visual design of the *Lecturer* condition, an informal home office was chosen – instead of a monochromatic green screen [32], together with a low text density [34]. For the usage of visual transitions, the definition of Kim et al. [18] was applied: Use of a primary visual representation for each of the treatment conditions without *major* changes that could affect the measurement of the dependent variables. Combining comprehension questions as recall assessment and application tasks as a transfer assessment has been done in previous studies, e.g., [20, 26, 43]. A previous comparison by Cross et al. [7] between a blended and an online class showed again no significant differences. Poquet et al. [30] analyzed the variables used in previous studies and found that many studies use video or presentation as independent variables. For the dependent variables, recall tests and transfer tests combined with learner self-reports are common.

2.4 Popular Public Formats

One format often highlighted is the "Khan Academy style", which uses a drawing board and voice-over-narration (e.g. in [12]). This approach can also be critically discussed, as Schwartz did [36], highlighting the pseudo-learn effect achieved through videos. In the context of his analysis, he uses the very popular Khan Academy videos. At no point in these videos is there a speaker visible, and the illusion of understanding is present. In the context of the "Khan style", a comparison between handwritten notes and *processed* handwritten notes has been conducted by Cross et al. [8] on major MOOC platforms, which revealed a preference for handwritten notes with post-processing. Another observation was summarized by Derek Muller, PhD and host of the popular YouTube channel

Veritasium [27]: *A great educational video allows you to leapfrog on someone else's thinking*. In the same video, he outlines that he has found a negative connection between learners' perceived ease of videos and their actual learning progress, which underlines the aforementioned illusion of understanding. Finally, on public video platforms, the face seems to be a relevant success factor for the viewer's decision which video to watch, e.g., in a thumbnail on YouTube. In non-academic blog posts, one finds the recommendation to include faces to increase the number of views [31]. The research community has used facial expressions and gestures to automate the generation and selection of these thumbnails [1, 37]. Shimono et al. summarize the analysis of YouTube thumbnails as follows "*We found that the facial expressions of the YouTuber in the frame are rich, the subject of the video is clear, and the content of the video is clearly defined by the text headline*". The learning items within a MOOC learning session do not compete with each other due to an existing order and linearity. Nevertheless, it can be argued that the visual appearance familiar to users from their private contexts influences their evaluation regarding the quality and visual experience in educational settings.

2.5 Research questions

Against this background, the following qualitative research question is asked, divided into three sub-questions: RQ1: Does the presence of an instructor influence a learner's perception regarding

1.1 the content?
1.2 the instructor?
1.3 his/her own learning?

The variable "perception" of a learner is operationalized through user feedback. Through a questionnaire containing twelve questions, these first three research questions are formed. Additionally, the given study aims to replicate previous research designs related to knowledge acquisition following the application of a treatment condition. Namely the recall and transfer assessment of a computer science concept. Therefore, the quantitative research questions are: RQ2: Does the presence of an instructor influences

1.1 the score of a recall assessment?
2.2 the score of a transfer assessment?

As prior research results vary regarding causal relationship between the video format and learner progress, as well as the relation between format and learner reception, the hypothesis for each RQ focuses on one measure of difference:

H0: The variable between the treatment conditions **is not** significantly different.
H1: The variable between the treatment conditions **is** significantly different.

2.6 Novelty of the study

Within the research community, the workshops at previous L@S-conferences, e.g., by Ritter et al. [33] in 2021, underlined the relevance of scaled A/B-tests and experiments. As the related work section shows, the presence of a lecturer leads to ambivalent results. Our study applies previous research designs to a scaled field experiment with adult learners. This is in contrast to smaller HCI experiments conducted on campus. In this way, we contribute four aspects to the ongoing debate:

- **Use of a high-stake exam of adult learners:** Learners can only participate once and within a certain time frame. Traditional classroom experiments are biased by university credit or low difficulty exams, and have lower internal validity.
- **Transparent and purposeful media production:** As mentioned before, previous studies have relied on existing learning material. Typically, the process of creation and publication is omitted as part of the research. By incorporating new video practices and providing access to the video material, we will broaden the discussion of effective audiovisual learning.
- **Focus on the domain of computer science:** While studies in biology [40], mathematics [43] and social sciences [20] have been used for comparison, to our best knowledge, the domain of computer science concepts and particularly an implementation has not been studied in terms of the impact of lecturer presence. Most of the articles do not specify the learning material used, so it is not possible to compare between lecture subjects.
- **Conducted after the Sars-CoV-2 pandemic:** The first comparisons began around 2000, followed by series of eye-tracking experiments (see the literature review of [30]). Now, after months of videoconferencing and digital collaboration, one could argue, that the perception and appeal of digital media has changed. Our data contextualize this potential effect to a large cohort of adult learners.

3 COURSE DESIGN

The research was conducted in a German speaking four-week MOOC[1], which was actively supervised by the research team. After these four weeks, the course content remains online in a self-paced course offering. All the data of the given study were taken from the actively supervised part of the course, meaning that only data from active participants were included. Future work might compare cohorts of the same course in different years or compare outcomes of the actively supervised course with those of self-paced course. The content of the course consisted of data structures and algorithms using Java. Solid basic Java knowledge in theory and practice was assumed as a prerequisite. The course reached this target group, as 90% of the users indicated that they have previously taken a Java programming course (on any platform). 84% of learners were previously enrolled in a course on our platform and therefore knew how to use the platform in terms of usability. To lower technical hurdles and to reduce the impact of confounding factors in our measurements, tutorials and introductory videos were offered, specifically targeting the 16% of first-time enrollments to the platform. Since there was not a single technical issue related to usability in the course forum, this is considered achieved. As the course was offered free of charge and with no enrollment restrictions – as the term open in MOOC suggests – there was no validation or testing prior to enrollment. The general openness of the participants to use video formats for learning can be described as high: First, the platform offers MOOCs throughout the year with a focus on

[1] https://open.hpi.de/courses/java-algorithmen2021

Table 1: Response rates of unique learner in terms of views, responses and submissions in the context of the experiment phases.

	Registration Phase and sub-total		Learners
	Begin of course		3,060
	Middle		3,437
	Course end		4,429
	Show-rate 66.33% = 2,938 unique learners		
	Experiment Phase and Participants		**Total n**
I	Pretest: Self-reported skill		1,280 submission
II	Treatment Lecturer (A): 356	Control condition Slides (B): 361	717 views
III	Survey (A): 245	Survey (B): 245	490 responses
IV	Posttest Recall (Quiz)		549 submission
V	Posttest Transfer (Programming Task)		690 submissions

information technology and digital skills. The general format and scheme of videos, quizzes and programming exercises are familiar to our learners, as is the optional use of the forum. Second, a survey (n=1,405) at the beginning of the course asked about personal preferences regarding learning media. A vast majority chose "Videos and animations" (79%), followed by "Text (digital access)" and "Pictures & Images" (both 62%) in this multi-select question. "Personal discussion" (30%) and "Podcasts and Audio-Books" (14%) were chosen least often.

88% of the users were based in Germany, followed by Austria (3.4%), the United States (2.64%) and Switzerland (2.23%). Similar to previous courses on our platform, the age structure represents adult learners (course average 45 years, platform average 42 years), who are primarily motivated by "*staying up to date*" on IT and CS topics. Of the 1,490 submissions regarding the primary goal of enrollment, 75% (n=1,118) responded "*I am interested in the topic*", while only 12% (n=179) sought a certificate of participation (CoP). With 855 CoPs, the number of certificates actually issued is higher than the self-reported goal (35% compared to 2,398 active learners at course end). The number of records of achievement is slightly lower (26%, n=643), due to the higher requirements: Learners had to achieve more than 50% of the total available points (quiz, programming task, homework). A breakdown of the response rates in the individual test phases is shown in Table 1.

In terms of device usage, a clear majority of 2,109 learners used the desktop web option, while 103 accessed the course exclusively via the mobile web, 83 users used the native apps for iOS and Android exclusively. A smaller proportion of users switched between desktop web and mobile web (163) or desktop web and native app (195).

4 METHODOLOGY

In order to ensure comparability between the two groups, a dedicated production process was associated with the research design and was conducted as follows.

4.1 Production Process

The lecture on recursion in Java was created specifically for this experiment. Half of the course participants was assigned to the *Lecturer* (A) and the other half to the *Slides* (B). The group allocation was randomized at the time of course enrollment (*round robin scheme*). In order to ensure comparability between the two treatment groups, a lecture script was written and discussed among the teaching team. The wording and the overall sentence structure used direct address to the audience (e.g. "*In this video, I would like to show you the basics of recursion using examples and pseudo code*") and informal presentation styles (e.g., by rhetorical questions: "*but... (pause) how does this work in a Java program?*"). Both type of elements are consistent with recommendations for the creation of educational YouTube videos [3]. The final version was then rehearsed for the recording session. The very same audio track was used for both versions of the learning material: While for the condition *Lecturer* the video and audio track of a speaker addressing the camera directly was used, for the condition *Slides* the audio track was used. A post-production editing was applied to both: For the slide condition, the PowerPoint slides were edited with the usual corporate identity design. Since this is the usual appearance of a learning item, the *Slides* treatment (group B) is considered as the control condition. For the *Lecturer* treatment, the same key elements were annotated on the video track. These annotations were always kept in line with the audio information to follow the redundancy principle as outlined by Moreno and Mayer [26]. The editing changes were carefully applied to both versions simultaneously in the editing software (Resolve Studio). As a result, both treatments show the same annotations (e.g., the definition of the base case, multimedia example of recursion), the same examples and the speaker uses exactly the same words (e.g., the example of the factorial 5!) at the same timestamp; therefore, both videos have a length of 4 minutes and 6 seconds. The Figure 1 above shows an exemplary frame at 1 minutes and 38 seconds. Before the final upload, the videos were tested by the teaching team, as well as independent academics which were less familiar with the topic. Both groups knew the lecturer (and his/her voice) from previous learning items, as the entire teaching team was introduced in an introductory segment prior to the first week. Since the two treatment groups have the same origin and identical processing steps, the technical details are also the same: Recorded in Full-HD (1080p), at 30 frames per second and a sample rate of 48kHz. Professional studio equipment was used for the recordings in an informal home office setting. Due to the natural, but informal setting of the *Lecturer* treatment, the split-attention effect should be lower because less attention is needed to focus on the slides and a lecturer. The same should apply to the slide condition, because the key elements of any given slide were highlighted in sync with the audio track. The video footage of the two video styles can be accessed here: http://las22.steinbeck.io.

4.2 Overview and Response Rates

All data presented refer to the period from November 24 to December 22 of 2021. Since the course was entirely voluntary and anonymous for the learners, different response rate can be reported. Of a total cohort size of 4,429 learners at the end, 2 out of 3 accessed the course at least once ('show-rate'). The six course announcements were read by an average of 1,588 people. A 90% response rate between submitted quiz questions and completed surveys can be

Table 2: Questions of the recall posttest

#	Questions and (Points)	Type
1	Evaluate the following statement: Basically, a task can be solved iteratively or recursively. (1)	True/False
2	Evaluate the following statement: Recursive approaches are generally faster than iterative approaches. (1)	True/False
3	Which line forms the "base case" of the recursion in the following code example? (2)	Single choice
4	The basic principle of dividing a large problem into many smaller problems is called... (1)	Single choice

considered high. The survey was neither rewarded nor mandatory. The set of 498 responses was balanced to include 245 responses in each treatment condition. The experiment proceeded in five phases, as shown in Table 1.

4.2.1 Pretest. In order to examine the existing knowledge on the topic of the lecture, a pretest was conducted in the form of a survey. Several main topics covered in the MOOC were queried and could only be submitted once. Thus, the field-experiment was not exposed and a comparison between course participants is possible. Two level of skills were surveyed. The first question, related to existing knowledge, used a 5-point ordinal scale, ranging from "*1- Have no knowledge*" to "*5 – I am confident to instruct others*". A second question referred to the existing ability of practical experience to program a recursive method (yes vs. no). The two questions therefor correspond to the two posttests in terms of depth of competence (recall and transfer).

4.2.2 Measuring perception – Survey. After watching their treatment group conditions, learners were asked to participate in a voluntary survey. The questionnaire was divided into three main parts – content, speaker and learner perspective. Table 3 summarizes the (translated) questionnaire and its results for measuring the perception variable. All questions were presented on a 5-point Likert-agreement-scale, except for L3 (multiple choice) and L4 (single choice). The survey was linked to a self-hosted LimeSurvey instance and the questionnaire was identical for both treatment groups.

4.2.3 Posttest 1 - Quiz. Although the videos could be (re-)viewed multiple times, each learner was only allowed to take the recall quiz once. The assessment was marked as a bonus-quiz, similar to a graded homework assignment with a due date. This deadline corresponds to the supervised portion of the course and the entire four-week time frame of the outlined study data. In conjunction with a time limit of 12 minutes, a high-stake exam was simulated. Table 2 shows the four quiz questions. Correctly identifying the base case in a given code (question three) was considered more valuable as it is a first step towards an implementation competency. The last question (*The basic principle of dividing a large problem into many smaller subproblems is called...*) serves as a proxy for measuring attention. The correct solution was explained in the very last part of each treatment and was not visualized or written out in any treatment. To get the correct answer (*divide and conquer*), learners had to pay attention throughout the video. Additionally, one survey question asked learners how much attention they paid ("*I watched the video carefully*"). A third component to measure attention was viewing time per item.

4.2.4 Posttest 2: Programming (Transfer). The implementation competence is assessed by a programming task. Within the virtual programming environment (Cloud IDE), each participant has the task of implementing a recursive function in Java. Since the course platform provided this web-based IDE, learners were able to focus on the task at hand. With this approach, technical issues and local setup problems can be eliminated as potentially demotivating factors and impediments. How-to knowledge of this online IDE was introduced at the beginning of the course and usability was ensured through prior tutorials and warm-up exercises. The intended solution resembled a Fibonacci sequence but was shifted by 1 to ensure that copied textbook solutions would not satisfy the automated unit tests. The assertions class and pattern matching were used to test two conditions: 1) Whether a for-loop – and thus an iterative and not recursive solution – was used and 2) whether the correct results were calculated. These tests ensured that only recursive solutions with correct calculation received a 100% score. Users were presented a read-only main class with function calls already prepared, which computed three instances [fib(2), fib(5) and fib(6)]. A second empty class was provided, for the actual assignment – writing the recursive solution to the given problem. A hidden fourth test case [fib(9)] was called in the background to avoid custom function calls for only three instances, but to test the learner's code for a general functionality.

5 RESULTS

5.0.1 Pretest. The majority reported having good or very good (4 and 5 out of 5) knowledge of recursion (n = 1,262). On the extreme ends, 14% (n=182) had no knowledge (1 out of 5) or felt confident to instruct others on the topic (highest option 5). 152 learners (12%) heard the term 'recursion' before (2/5), 279 (22%) report a "slight idea" (3/5), leaving 469 learners who believe they understood the topic (level 4 out of 5). In the second pretest question, a majority said they already had practical programming experience with recursion (711 yes; 551 no).

5.0.2 Video statistics. The average viewing rate of the 42 videos in the course was 98.6%. *Lecturer* had an average farthest viewing time of 97%, the *Slide* group watched an average of 100%; For treatment A (*Lecturer*), 118 forward and 389 backward calls were counted, while for condition B *Slide* 122 forward and 438 backward calls were recorded. The total number of unique viewers for A was 359, while for condition B it was 368 views. Based on these indicators, no difference in the use of the video elements could be detected.

5.0.3 Perception – Survey. The twelve questions were tested with a non-parametric test (Mann-Whitney U, pairwise deletion in SPSS v23), due to their ordinal nature and non-normal distribution (confirmed by Shapiro-Wilk test). This test was chosen to find significant difference between the treatment groups and thus to reject or retain the outlined hypotheses. Overall, the two groups show very similar responses in terms of perception of the content, the instructor and

Table 3: Survey results of C (Content), S (Speaker) and Learner (L), means on a 5-Likert agreement-scale (1-No agreement to 5-Total agreement); L4 on a single choice scale between 1 (grade A) and 6 (grade F); *significantly different at p<0.01 tested by a Mann-Whitney U test, grouped by experiment conditions (*Lecturer* vs. *Slides*).

		Lecturer				*Slides*			
#	Long Question Text	n	Mode (n)	Mean	SD	n	Mode (n)	Mean	SD
C1	The content presented was easy to understand.	237	5 (119)	4.36	.799	239	4 (108)	4.22	.858
C2	The principle of recursion was presented in a visually clear way.	238	5 (109)	4.23	.910	236	5 (99)	4.17	.899
C3	The length of the video explained the topic well.	235	5 (135)	4.38	.876	231	5 (109)	4.26	.891
C4	The content is the same as what I am used to on openHPI.	230	5 (102)	4.06	1.072	225	5 (100)	4.13	1.00
S1	The speaker made a professional impression.	232	5 (121)	4.38	.807	235	5 (117)	4.38	.755
S2	I had the feeling that the speaker was in the same room as me.	218	3 (66)	3.14	1.16	222	3 (81)	3.27	1.10
S3	I felt like the speaker was focusing on me.	221	4 (80)	3.58*	1.07	217	3 (81)	3.28*	1.12
S4	I felt like the speaker knew I was watching the video to learn.	218	5 (92)	4.07	1.02	220	5 (81)	3.88	1.10
L1	I watched the video attentively.	235	5 (135)	4.43	.836	234	5 (119)	4.38	.751
L2	I could easily follow what was said.	232	5 (126)	4.37	.868	234	5 (118)	4.34	.809
L3	I think seeing the instructor in the video is, ___ compared to not seeing him.	See descriptive Figure 2							
L4	What overall grade would you give the video?	236	2 (123)	1.87*	.752	235	2 (131)	2.09*	.857

their own learning. C1, C2 and C4, S1 and S2, as well as L1 and L2 clearly provide similar results. C3 and S4 are on the edge of a significant difference. The perceived level of professionalism (S1) is exactly the same. This may highlight the importance of audio over video as described in the literature. The similarity of C4 is surprising, as the *Lecturer* video style is significantly different from the usual visual experience of the platform, while *Slides* in the corporate design fit more the status quo. Additionally, the perception of length (C3) is an interesting candidate for further research, since both video styles were of equal length. This could indicate that the visual experience influences the perceived duration of a learning item, as the *Slides* treatment has a lower correspondence. Further research could investigate this relationship and analyze the range of length at which seeing a lecturer can increase attentively followed lecture time.

A significant different was measured for the attention of the speaker (S3) (U: 20,219; z-value: -2.94; p<0.01). The difference could only be derived from the visual experience, as *Lecturer* looked into the camera and addressing the learner directly. The other significant perception in this study is the overall grade (L4) given by the learners. On a scale between "very good" (1) and "deficient" (6), the *Lecturer* treatment shows a higher grade (1.87) than the control condition (2.09) (U: 23,896.5; z-value: -2.86; p<0.01).

Although L3 in the form of a projection question showed no statistical difference, the descriptive analysis replicates the findings by Ng and Przybyłek [28]: The adjectives with positive connotations *useful*, *helpful* and *pleasant* were chosen eleven-times more often than their negative counterparts (685 to 61). The neutral statement *it does not matter* was chosen 83 times in total and slightly more often in the *Slides* group as Figure 2 shows.

Consequently, the null hypotheses for the research questions 1.1, 1.2 and 1.3 can be **retained, hence rejecting H1**: The presence of a lecturer shows **no** statistically significant difference in learners' perception of the content, the instructor and their own learning.

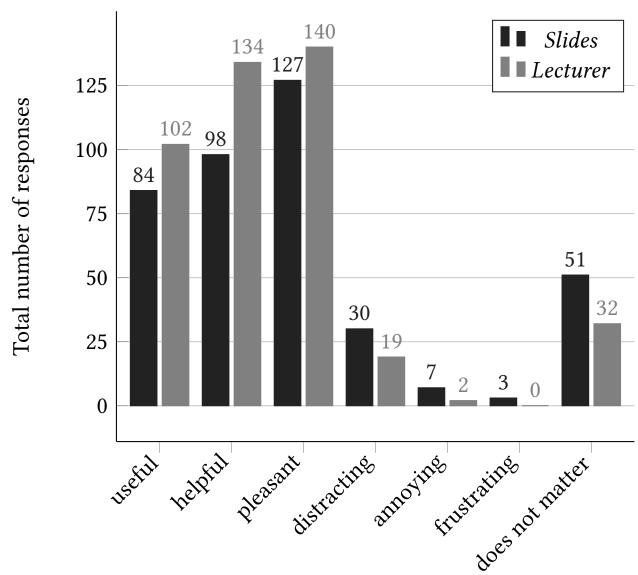

Figure 2: Positive and negative attributes regarding seeing the lecturer, multiple answer choice.

5.0.4 Posttest 1 – Quiz (Recall). Since the mode of the pretest was 4 and the average was 3.25, a similarly high score on the recall assessment could be expected. With an average of 3.87 points (equals to 77%), the average quiz performance is higher than the average self-reported skill level within the overall cohort. Compared to the overall quiz performance in the course (61%), the recall test scores about recursion are higher. Similar to the pretest, a Shapiro-Wilk

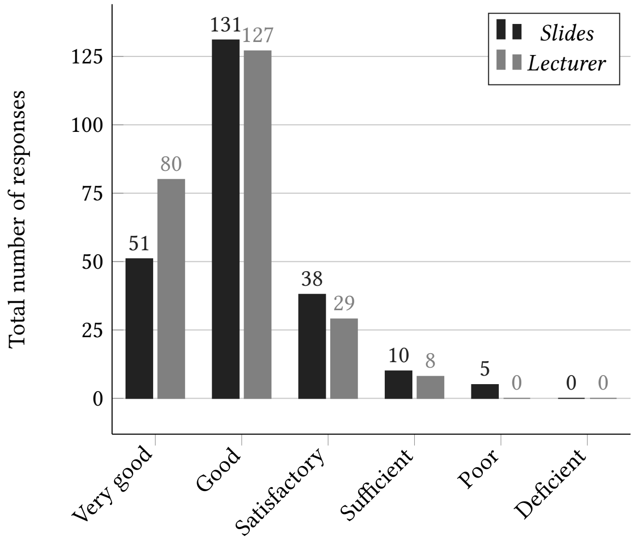

Figure 3: Overall quality perception, measured by grades (1 - very good, to 6 - inadequate.

test for normality revealed non-normally distributed results. A histogram report confirmed that as well. The following results show the statistical analysis with a t-test[2], resulting in a statistically significant difference between treatment groups with a small effect size (t(547)= -2.248, p< .05, Cohens d= .191). Additionally, the *Slides* group needed significant more time than the *Lecturer* treatment group (t(547)= 2.491, p< .05). In order to analyze the point difference between the two groups, the average treatment effect (ATE) (see[15] and [2]) was calculated using Stata. On a scale between 0 and 1, the coefficient between *Lecturer* versus *Slides* yields a 0.264 (z= 2.22; p< 0.05), thus explaining the average difference between the groups with approximately a quarter point. Although relatively small, a quarter point represents 5% worth in the recall posttest and underlines the statistical difference with a small effect size between the groups.

Table 4: Overview of the quiz score and timings of learners who submitted the recall post-test *and* submitted the survey.

Descriptive Statistics		by Quiz Score		by Quiz Duration	
Experiment Group	n	Mean	SD	Mean	SD
Slides	279	3.76	1.16	189 sec	139 sec
Lecturer	270	3.98	1.15	161 sec	127 sec

Consequently, the null hypotheses for RQ 2.1 can be **rejected, hence accepting H1**: The presence of a lecturer shows a statistically significant different in the **recall** assessment.

[2]Although the test is not applicable at first glance, it can be considered robust in large samples due to the non-normal distribution, as shown by Edgell and Noon [10] and Lumley et al.[22]

5.0.5 Posttest 2 – Programming (Transfer). Of the 690 unique submissions, 657 received a 100% on the programming assignment, the remaining 33 learners received one out of two points. With a completion rate of 86.1%, an average score of 1.95/2 (97.6%) and an average completion task of approximately 8 min, the task is on par with the other programming exercises in the course and was neither too easy nor too hard to solve. Especially the high numbers of successful implementation of the recursive algorithm and the correct calculations indicate a high translation between the learning process and the application of knowledge. At the same time, the results fit the high self-reported skill-levels in the cohort. No statistically significant different can be reported between the two treatment groups.

Consequently, the null hypotheses for RQ 2.2 can be **retained (accepting H0)**: The presence of a lecturer shows **no** statistically significant difference in the **transfer assessment**.

6 DISCUSSION

This study compares two video styles in a pretest-posttest research design. In a MOOC of 2,938 active participants, the treatment group *Lecturer* viewed a modern talking-head format, while the control group (*Slides*) received a traditional PowerPoint presentation with a voice-over. A self-reported skill level regarding the concept of recursion was used as a pretest, followed by a quiz-based posttest; both on a five-point scale. Additionally, a second posttest focusing on transfer knowledge was integrated. While the posttest focusing on recall-knowledge shows a significant difference, the transfer knowledge does not. Overall, learners perceptions do not vary significantly between the two groups. The areas that show significant differences relate to the speaker's attention (S3 - "*I felt like the speaker was focusing on me*"), which replicates existing research findings on social presence in digital learning environments. Due to the carefully produced learning material, this difference is caused by the visual experience. Teaching into a camera and addressing learners directly leads to a stronger effect in terms of perceived attention of a speaker. The same result cannot be achieved by the same narration on slides, even with guided step–by–step bullet points as seen in many university lectures and MOOCs. A second impact caused by the video style is the overall quality assessment. Learners reported better qualitative grades in the *Lecturer* treatment group. As the *Lecturer* condition was designed and produced to be more visually appealing, the results can be interpreted in favor of showing a lecturer's face: It promotes presence and increases perceived learner quality. Combined with slightly higher scores in the recall task, we still argue in favor of showing a face in introduction classes and learning situations focusing on the acquisition of fact-knowledge. At the same time, the study results add to the existing amount of previous research that find little to none quantitative evidence in favor of a lecturer presence. As outlined in the beginning, several months of video-conferencing and consumption of video-based learning items does not change that evidence. In light of these replication findings, we recommend to shift the discussion to the dedicated production process: A script was written, revised and rehearsed specifically for online teaching environments. The script then serves as an supporting media that can be presented with a person, slides, animations, programming code or drawing tablet-style. Here, the presentation medium is less important than the underlying, "supporting" media. Especially the results of the transfer skill underline this, as both groups were able to solve the implementation without measurable difference. The similarity of S1 ("*The speaker made a professional impression*") can be caused by the usage of professional equipment. The literature emphasizes the importance of clear audio over good video quality. Since the audio quality was the same in both groups, we assume that voice still has an important influence of the perceived professionalism of a learning item. Surprisingly, the responses to C4 ("*The content is the same as what I am used to on openHPI*") are similar. Typically, the platform provides a picture-in-picture view of slides along with a speaker view in front of a monochromatic studio green screen. Thus, a difference in both groups were expected. Since this carries the risk of the split-attention-effect and a controlled measurement of attention is not possible, two different video styles were chosen. Neither the informal recording setting nor the traditional presentation resulted in an immersive perception of the physical space: Perception of physical presence (S2) showed lower agreement scores in both groups. In particular, since the mode represents „*neither agree nor disagree*", the effect on imitation of physical presence cannot be evaluated.

6.1 Limitations

The numbers of participants in a MOOC build a strong case of an underlying cause-and-effect relationship due to the natural learning environment. At the same time, each individual learner cannot be "captured" in detail with respect to the experiment variables. As previous research has shown, perceived presence in particular, operationalized through a user-survey, is subject to bias and can only be used as an approximation [28]. Although server-timings and access logs supported the study, no holistic view of a user's learning behavior can be derived: *What is a learner doing in a second browser tab? How is additional information being accessed?* Second, the results of this study represents primarily adult learners between the ages of 30 and 50. External validity can only relate to this specific age group, even though there are younger and older learners in the cohort. Differences in demographics and especially age could be explored in further studies to determine if a full-time student or the "Tik Tok-generation" would rate the learning materials differently. As the literature shows, these universal video platforms use thumbnails, clickbait content and a lead story differently then traditional lectures.

7 CONCLUSION

The correct measures of "humanizing" and create appealing learning videos in digital environments is an ongoing debate. Whether it's in a video conferencing session with a wall of inactive (students) cameras or a lecturer teaching into a camera. In scaling learning environments, asynchronous distribution of video material that can be accessed on demand is the common proposal. The initial MOOC phase used existing learning environments in universities, followed by a phase in which existing slides from the same university context were recycled and put into a different time-frame than the traditional 90-minute lecture. Creating specific learning content for a dedicated course initiated the third phase. In light of this increasing maturity, this study contributes to the debate on effective instruction at scale. In summary, given the momentum and acceptance of audiovisual media, the influence of the scientific community can shape the production process of MOOCs and education videos in general. This includes the adaptation of popular video formats from general video platforms (e.g., YouTube, TikTok, Twitch) and the commitment to analyze these formats and platforms in terms of learning effectiveness and perceived usefulness. The resulting debate helps the higher education community - both researchers and lecturers – to evaluate whether it is about slides, faces, gestures or an overall compelling educational narrative. In order to facilitate this debate, we advocate for more data points, particularly in the form of longitudinal comparisons across multiple learning items and course weeks. Although costly and complex, a cohort comparison of the same MOOC with different audiovisual elements could yield interesting results. Together with the outlined research findings from closely monitored eye-tracking experiments, a holistic view of modern video formats in higher education emerges.

REFERENCES

[1] Evlampios Apostolidis, Eleni Adamantidou, Vasileios Mezaris, and Ioannis Patras. 2021. Combining Adversarial and Reinforcement Learning for Video Thumbnail Selection. In *Proceedings of the 2021 International Conference on Multimedia Retrieval*. ACM, Taipei Taiwan, 1–9. https://doi.org/10.1145/3460426.3463630
[2] Orley Ashenfelter. 1978. Estimating the Effect of Training Programs on Earnings. *The Review of Economics and Statistics* 60, 1 (1978), 47–57. https://doi.org/10.2307/1924332 Publisher: The MIT Press.
[3] Jacob Beautemps and André Bresges. 2021. What Comprises a Successful Educational Science YouTube Video? A Five-Thousand User Survey on Viewing Behaviors and Self-Perceived Importance of Various Variables Controlled by Content Creators. *Frontiers in Communication* 5 (2021). https://www.frontiersin.org/article/10.3389/fcomm.2020.600595
[4] Maik Beege, Sascha Schneider, Steve Nebel, and Günter Daniel Rey. 2017. Look into my eyes! Exploring the effect of addressing in educational videos. *Learning and Instruction* 49 (June 2017), 113–120. https://doi.org/10.1016/j.learninstruc.2017.01.004
[5] Jianxia Cao, Akinori Nishihara, and Shijuan Wang. 2018. Effect of instructor inclusion type and course complexity on students' learning in lecturer video. In *Proceedings of the 10th International Conference on Education Technology and Computers - ICETC '18*. ACM Press, Tokyo, Japan, 181–188. https://doi.org/10.1145/3290511.3290544
[6] Tiphaine Colliot and Éric Jamet. 2018. Understanding the effects of a teacher video on learning from a multimedia document: an eye-tracking study. *Educational Technology Research and Development* 66, 6 (Dec. 2018), 1415–1433. https://doi.org/10.1007/s11423-018-9594-x
[7] Andrew Cross, B. Ashok, Srinath Bala, Edward Cutrell, Naren Datha, Rahul Kumar, Viraj Kumar, Madhusudan Parthasarathy, Siddharth Prakash, Sriram Rajamani, Satish Sangameswaran, Deepika Sharma, and William Thies. 2014. Online learning versus blended learning: an exploratory study. In *Proceedings of the first ACM conference on Learning @ scale conference*. ACM, Atlanta Georgia USA, 179–180. https://doi.org/10.1145/2556325.2567869
[8] Andrew Cross, Mydhili Bayyapunedi, Edward Cutrell, Anant Agarwal, and William Thies. 2013. TypeRighting: combining the benefits of handwriting and typeface in online educational videos. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '13)*. Association for Computing Machinery, New York, NY, USA, 793–796. https://doi.org/10.1145/2470654.2470766
[9] Brian Dean. 2021. How Many People Use YouTube in 2022? Retrieved January 4, 2022 from https://backlinko.com/youtube-users
[10] Stephen E. Edgell and Sheila M. Noon. 1984. Effect of violation of normality on the t test of the correlation coefficient. *Psychological Bulletin* 95, 3 (1984), 576–583. https://doi.org/10.1037/0033-2909.95.3.576
[11] Charlotte N. Gunawardena. 1995. Social Presence Theory and Implications for Interaction and Collaborative Learning in Computer Conferences. *International Journal of Educational Telecommunications* 1, 2 (1995), 147–166. https://www.learntechlib.org/p/15156 Place: Charlottesville, VA Publisher: Association for the Advancement of Computing in Education (AACE).

[12] Philip J. Guo, Juho Kim, and Rob Rubin. 2014. How video production affects student engagement: an empirical study of MOOC videos. In *Proceedings of the first ACM conference on Learning @ scale conference*. ACM, Atlanta Georgia USA, 41–50. https://doi.org/10.1145/2556325.2566239

[13] Anna Hansch, Lisa Hillers, Katherine McConachie, Christopher Newman, Thomas Schildhauer, and J. Philipp Schmidt. 2015. *Video and Online Learning: Critical Reflections and Findings from the Field*. SSRN Scholarly Paper ID 2577882. Social Science Research Network, Rochester, NY. https://doi.org/10.2139/ssrn.2577882

[14] Jennifer S Hegeman. 2015. Using Instructor-Generated Video Lectures in Online Mathematics Courses Improves Student Learning. 19, 3 (2015), 18.

[15] Paul W. Holland. 1986. Statistics and Causal Inference. *J. Amer. Statist. Assoc.* 81, 396 (1986), 945–960. https://doi.org/10.1080/01621459.1986.10478354 Publisher: Taylor & Francis _eprint: https://www.tandfonline.com/doi/pdf/10.1080/01621459.1986.10478354.

[16] Bruce D. Homer, Jan L. Plass, and Linda Blake. 2008. The effects of video on cognitive load and social presence in multimedia-learning. *Computers in Human Behavior* 24, 3 (May 2008), 786–797. https://doi.org/10.1016/j.chb.2007.02.009

[17] Slava Kalyuga, Paul Chandler, and John Sweller. 1998. Levels of Expertise and Instructional Design. *Human Factors* 40, 1 (1998), 1–17. https://doi.org/10.1518/001872098779480587 Publisher: SAGE Publications Inc.

[18] Juho Kim, Philip J. Guo, Daniel T. Seaton, Piotr Mitros, Krzysztof Z. Gajos, and Robert C. Miller. 2014. Understanding in-video dropouts and interaction peaks inonline lecture videos. In *Proceedings of the first ACM conference on Learning @ scale conference (L@S '14)*. Association for Computing Machinery, New York, NY, USA, 31–40. https://doi.org/10.1145/2556325.2566237

[19] René F. Kizilcec, Jeremy N. Bailenson, and Charles J. Gomez. 2015. The instructor's face in video instruction: Evidence from two large-scale field studies. *Journal of Educational Psychology* 107, 3 (2015), 724–739. https://doi.org/10.1037/edu0000013

[20] René F. Kizilcec, Kathryn Papadopoulos, and Lalida Sritanyaratana. 2014. Showing face in video instruction: effects on information retention, visual attention, and affect. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, Toronto Ontario Canada, 2095–2102. https://doi.org/10.1145/2556288.2557207

[21] Mehmet Kokoç, Hale Ilgaz, and Arif Altun. 2020. Effects of sustained attention and video lecture types on learning performances. *Educational Technology Research and Development* 68, 6 (Dec. 2020), 3015–3039. https://doi.org/10.1007/s11423-020-09829-7

[22] Thomas Lumley, Paula Diehr, Scott Emerson, and Lu Chen. 2002. The Importance of the Normality Assumption in Large Public Health Data Sets. *Annual Review of Public Health* 23, 1 (May 2002), 151–169. https://doi.org/10.1146/annurev.publhealth.23.100901.140546

[23] Jeff Maggioncalda. 2021. *Building a more just world: The role of institutions in defining the new normal.* Retrieved December 12, 2021 from https://open.hpi.de/courses/hpi-learningatscale2021-emoocs2021/items/1qgfCzypxxxzO6X7JKVXf8

[24] Richard E. Mayer, Julie Heiser, and Steve Lonn. 2001. Cognitive constraints on multimedia learning: When presenting more material results in less understanding. *Journal of Educational Psychology* 93, 1 (2001), 187–198. https://doi.org/10.1037/0022-0663.93.1.187

[25] Roxana Moreno and Richard E Mayer. 1999. Cognitive Principles of Multimedia Learning: The Role of Modality and Contiguity. (1999), 11.

[26] Roxana Moreno and Richard E. Mayer. 2002. Verbal redundancy in multimedia learning: When reading helps listening. *Journal of Educational Psychology* 94, 1 (2002), 156–163. https://doi.org/10.1037/0022-0663.94.1.156

[27] Muller, Derek Veritasium. 2017. How Should We Teach Science? Retrieved January 5, 2022 from https://youtu.be/Gstcvew6FVM?t=30

[28] Yen Ying Ng and Adam Przybyłek. 2021. Instructor Presence in Video Lectures: Preliminary Findings From an Online Experiment. *IEEE Access* 9 (2021), 36485–36499. https://doi.org/10.1109/ACCESS.2021.3058735 Conference Name: IEEE Access.

[29] Cuong Nguyen and Feng Liu. 2016. Gaze-based Notetaking for Learning from Lecture Videos. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*. ACM, San Jose California USA, 2093–2097. https://doi.org/10.1145/2858036.2858137

[30] Oleksandra Poquet, Lisa Lim, Negin Mirriahi, and Shane Dawson. 2018. Video and learning: a systematic review (2007–2017). In *Proceedings of the 8th International Conference on Learning Analytics and Knowledge (LAK '18)*. Association for Computing Machinery, New York, NY, USA, 151–160. https://doi.org/10.1145/3170358.3170376

[31] Renderforest. 2021. 13 Tips for a Clickable YouTube Video Thumbnail. Retrieved January 7, 2022 from https://www.renderforest.com/blog/youtube-video-thumbnail-tips#scroll-nav__3

[32] Jeanine Reutemann. 2016. Differences and Commonalities – A comparative report of video styles and course descriptions on edX, Coursera, Futurelearn and Iversity. In *Proceedings of the 4th European MOOCs Stakeholders Summit (EMOOCs 2016)*. 383–392.

[33] Steve Ritter, Neil Heffernan, Joseph Jay Williams, Derek Lomas, and Klinton Bicknell. 2021. Second Workshop on Educational A/B Testing at Scale. In *Proceedings of the Eighth ACM Conference on Learning @ Scale* (Virtual Event, Germany) *(L@S '21)*. Association for Computing Machinery, New York, NY, USA, 1–3. https://doi.org/10.1145/3430895.3460876

[34] José Miguel Santos Espino, M.D. Afonso-Suárez, and Cayetano Guerra Artal. 2016. Speakers and boards: A survey of instructional video styles in MOOCs. 63 (05 2016), 101–115.

[35] José Miguel Santos Espino, María Dolores Afonso Suárez, and Juan José González-Henríquez. 2020. Video for teaching: classroom use, instructor self-production and teachers' preferences in presentation format. *Technology, Pedagogy and Education* 29, 2 (March 2020), 147–162. https://doi.org/10.1080/1475939X.2020.1726805

[36] Marc Schwartz. 2013. Khan Academy: The Illusion of Understanding. *Online Learning Journal* 17, 4 (Dec. 2013). https://www.learntechlib.org/p/183760/

[37] Akari Shimono, Yuki Kakui, and Toshihiko Yamasaki. 2020. Automatic YouTube-Thumbnail Generation and Its Evaluation. In *Proceedings of the 2020 Joint Workshop on Multimedia Artworks Analysis and Attractiveness Computing in Multimedia*. ACM, Dublin Ireland, 25–30. https://doi.org/10.1145/3379173.3393711

[38] Andrew T. Stull, Logan Fiorella, and Richard E. Mayer. 2018. An eye-tracking analysis of instructor presence in video lectures. *Computers in Human Behavior* 88 (Nov. 2018), 263–272. https://doi.org/10.1016/j.chb.2018.07.019

[39] Sullygnome.com. 2021. Twitch stats and analysis - for the year 2021. Retrieved January 4, 2022 from https://sullygnome.com/2021

[40] Sera Thornton, Ceri Riley, and Mary Ellen Wiltrout. 2017. Criteria for Video Engagement in a Biology MOOC. In *Proceedings of the Fourth (2017) ACM Conference on Learning @ Scale*. ACM, Cambridge Massachusetts USA, 291–294. https://doi.org/10.1145/3051457.3054007

[41] Judith Uchidiuno, Jessica Hammer, Evelyn Yarzebinski, Kenneth R. Koedinger, and Amy Ogan. 2017. Characterizing ELL Students' Behavior During MOOC Videos Using Content Type. In *Proceedings of the Fourth (2017) ACM Conference on Learning @ Scale*. ACM, Cambridge Massachusetts USA, 185–188. https://doi.org/10.1145/3051457.3053981

[42] Margot van Wermeskerken and Tamara van Gog. 2017. Seeing the instructor's face and gaze in demonstration video examples affects attention allocation but not learning. *Computers & Education* 113 (2017), 98–107. https://doi.org/10.1016/j.compedu.2017.05.013

[43] Jiahui Wang and Pavlo D. Antonenko. 2017. Instructor presence in instructional video: Effects on visual attention, recall, and perceived learning. *Computers in Human Behavior* 71 (June 2017), 79–89. https://doi.org/10.1016/j.chb.2017.01.049

[44] Aimee L. Whiteside, Amy Garrett Dikkers, and Karen Swan. 2017. *Social Presence in Online Learning: Multiple Perspectives on Practice and Research.* Stylus Publishing, LLC.