Distance between gaze and laser pointer predicts performance in video-based e-learning independent of the presence of an on-screen instructor

Marian Sauter
Ulm University, Institute of Psychology and Education, General Psychology
sauter.marian@gmail.com

Tobias Wagner
Ulm University, Institute of Media Informatics
tobias.wagner@uni-ulm.de

Anke Huckauf
Ulm University, Institute of Psychology and Education, General Psychology
anke.huckauf@uni-ulm.de

ABSTRACT

In online lectures, showing an on-screen instructor gained popularity amidst the Covid-19 pandemic. However, evidence in favor of this is mixed: they draw attention and may distract from the content. In contrast, using signaling (e.g., with a digital pointer) provides known benefits for learners. But effects of signaling were only researched in absence of an on-screen instructor. In the present explorative study, we investigated effects of an on-screen instructor on the division of learners´ attention; specifically, on following a digital pointer signal with their gaze. The presence of an instructor led to an increased number of fixations in the presenter area. This did neither affect learning outcomes nor gaze patterns following the pointer. The average distance between the learner's gaze and the pointer position predicts the student's quiz performance, independent of the presence of an on-screen instructor. This can also help in creating automated immediate-feedback systems for educational videos.

CCS CONCEPTS

• **Human-centered computing**; • **Human computer interaction (HCI)**; • **Empirical studies in HCI**;

KEYWORDS

eye tracking, regression analysis, e-learning, test performance

ACM Reference Format:
Marian Sauter, Tobias Wagner, and Anke Huckauf. 2022. Distance between gaze and laser pointer predicts performance in video-based e-learning independent of the presence of an on-screen instructor. In *2022 Symposium on Eye Tracking Research and Applications (ETRA '22), June 08–11, 2022, Seattle, WA, USA.* ACM, New York, NY, USA, 10 pages. https://doi.org/10.1145/3517031.3529620

1 INTRODUCTION

The presence of an on-screen instructor is an uprising trend in educational video content [Henderson and Schroeder 2021], typically shown on the right side (either upper right or lower right corner). A recent systematic review summarized the impact on-screen instructors have on the learning outcomes (and affective state) [Henderson and Schroeder 2021]. While participants seemed to be more satisfied with the video when an on-screen instructor was present, the evidence towards a beneficial effect with regards to learning outcomes is mixed ([Henderson and Schroeder 2021], Table 1). While one study found a beneficial effect in an easy condition [Wang and Antonenko 2017], another one did not [Wang et al. 2020]. The latter study found a beneficial effect for a difficult condition, but only in a knowledge transfer test [Wang et al. 2020]. Another study found a positive effect explicitly for declarative knowledge, but not for procedural knowledge [Hong et al. 2018]. Most studies were limited to null findings but no study found a detrimental effect of an on-screen instructor regarding the learning outcomes (but it may, for example, increase cognitive load [van Wermeskerken et al. 2018]). Notably, some students reported that the on-screen instructor may have hindered their ability to focus properly [Henderson and Schroeder 2021; Yu 2021]. This is plausible, as faces, especially videos of humans, attract eye movements and capture attention [Bindemann et al. 2005; Langton et al. 2008; Wang et al. 2020]. It stands to reason that only those participants, who are not too distracted by the on-screen instructor and coherently focus on the relevant part of the slide can benefit (cognitively or affectively, resulting in improved learning outcomes) from the instructor presence.

Lecturers often try to direct the learner's focus to the important part of the slide (i.e. signaling), by means of hand gestures, a physical laser pointer or a digital laser pointer (the latter being most prominent in e-learning) [Knoblauch 2008; Pi et al. 2019]. Such directing of attention was shown to be generally beneficial for the learning of students [Badman et al. 2016; Pi et al. 2019; Sung and Mayer 2012] and has positive affective effects [Sung and Mayer 2012]. A study using the student's and lecturer's gaze as means of signaling (each other) found that pointing is generally perceived as helpful for students and lecturers alike [Špakov et al. 2019]. A recent study did not employ signaling but investigated the audience's eye movements when viewing short educational clips [Madsen et al. 2021]. They found that synchronicity of eye movements is predictive of test scores in a subsequent quiz. This means that participants who had a similar eye movement pattern compared to the 'average' participant, performed better. Supposing that on average, the participants attended to the most informative parts of the lecture slides, it seems plausible that they also would have attended and benefited from a signal (if it had been there). Another study [Sharma 2015] directly investigated how participants

ETRA '22, June 08–11, 2022, Seattle, WA, USA

ACM ISBN 978-1-4503-9252-5/22/06.
https://doi.org/10.1145/3517031.3529620

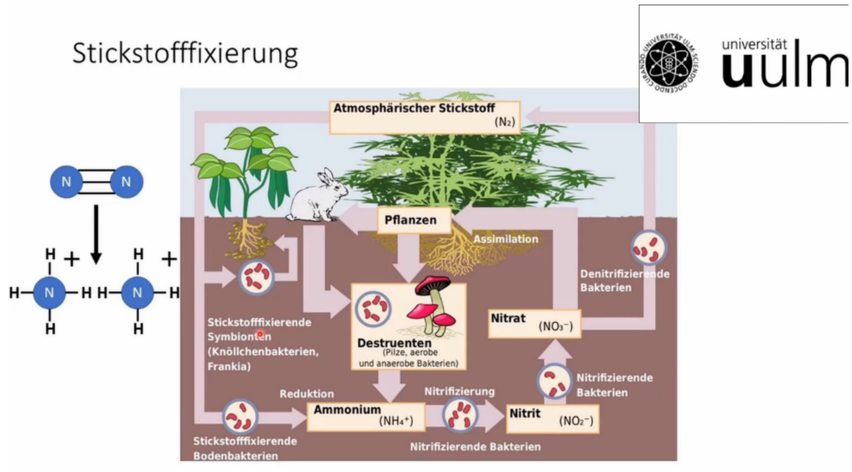

Figure 1: Example of a learning material slide (topic: nitrogen) including the box for the logo. For videos with a presenter video feed instead of a logo, the feed occupied the whole boxed area in the upper right corner.

follow verbal directions to specific parts of the content and how this relates to subsequent test performance. They found that, in fact, both the first fixation time (i.e. how long did participants first look at the verbally referenced site) and the refixation rate (i.e. how often do participants look back to the referenced site after looking elsewhere) were both positively correlated with the post-test performance. What the aforementioned studies have in common is that there is no lecturer visible in the learning videos (or the authors do not disclose whether there is one). Given the on-screen instructor's potential to capture attention and divert focus, it is unclear, whether the results from the signaling studies can be directly translated to settings with an on-screen instructor.

Therefore, in the present study, we explore whether participants are generally able to follow a digital laser pointer signal with their gaze in conditions with and without and on-screen instructor. In order to do this, we record eye movements of participants watching educational video clips. In particular, we investigate (1) how well the average distance between the signal position and the gaze position is predictive of subsequent performance in a quiz, (2) whether this relationship is affected by the presence of an on-screen instructor and (3) whether an on-screen instructor is beneficial or detrimental to post-test performance.

2 METHODS

2.1 Participants

Twenty-three local naïve psychology students (M_{Age} = 20 y, $Range_{Age}$ = 18 – 27 y; 8 male, 15 female) participated in our study. All had normal vision (glasses were an exclusion criterium due to eye tracking issues) and provided written informed consent. Twenty-one students indicated their highest degree as Abitur (local high school diploma) and two as a Bachelor's degree. Twenty-two participants indicated that the language of the study (German) was their native language and one participant indicated that they speak it fluently. They were compensated with 8€ or course credit. The study was in line with the ethics guidelines of Ulm University and conducted in line with the Declaration of Helsinki.

2.2 Video Material

Four short educational videos were created in a diagram-based style and covering topics that were not part of the psychology curriculum. The topics were (1) how a nuclear power plant works, (2) the nitrogen cycle, (3) how hurricanes form and (4) electrolysis. All videos were recorded with signaling using Microsoft PowerPoint (i.e. a virtual laser pointer) and the video feed of a presenter in the upper right corner. The pointing with the virtual laser pointer was achieved using PowerPoint's built-in function. The videos were between 03:51 minutes and 05:14 minutes long and consisted of two to five slides each. For investigating the hypotheses that a visible presenter has influence on eye movements, we created another version of all videos in which we covered the presenter with a static logo image (see Figure 1).

2.3 Experimental procedure and design

After an explanation of the procedure (for a visualization, see Figure 2) by the experimenter, the eye tracker was calibrated (see below) and participants started the experiment with a brief introductory questionnaire on demographic variables. In particular, the participants were asked to indicate their age, gender, proficiency of the experimental language and highest educational degree earned.

Afterwards, the first video started. When the video ended, participants were asked to indicate (1) how new the content was to them on a scale from 1 (everything new) to 5 (all well-known), (2) how difficult they experienced the video on a 5-point scale from very easy to very hard, and (3) how they liked the form of the presentation on a 5-point scale from very much to not at all. Afterwards, they answer four quiz questions regarding the video's content (see). This was repeated for all four videos. After the last quiz was answered, a post-questionnaire followed.

In the post-questionnaire, the participants were first asked to create an order regarding the difficulty of the videos by putting in

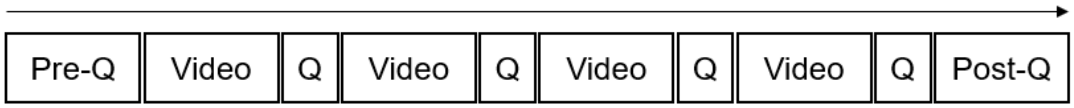

Figure 2: Diagram of the experimental procedure. After an introductory questionnaire (Pre-Q), participants watched four videos, each followed by a specific quiz on the content of the video (Q). At the end, another questionnaire followed (Post-Q).

the numbers from 1 to 4 next to the video's title, and then indicate how difficult it was to create this order. They further had to indicate how long they thought they relatively looked at the presenter/signal vs anywhere else by adjusting a slider (from left 'presenter/signal' to right 'anywhere else') and then how long they thought they relatively looked at the presenter vs the signal using another slider (from left 'presenter to right 'signal'). Concludingly, we asked a qualitative free-text question concerning their impression of the presenter's insertion or use of the laser pointer.

Each participant watched four videos covering all four educational topics. For each participant, two of them were presented in the version with the presenter's video feed and two of them were presented in the version with the static image (counterbalanced across participants).

2.4 Set up and eye movement recording

The experiment was built and conducted using the software OpenSesame 3.3.10 [Mathôt et al. 2012; OpenSesame 2022] using a Python 3.7.6. [Python.org 2022] backend. It was presented on a 27" screen at a resolution of 1920x1080 with the color scheme set to sRGB at a distance of 70cm. Participants indicated their responses using a regular two-button mouse or a regular keyboard, respectively.

For recording the participant's eye parameters, we used the desktop-mounted SR Research Eyelink 1000 Plus eye-tracker [SR Research 2021]. We further used the standard settings from the SR Research Eyelink built in the plugin version 0.3 for OpenSesame. We tracked the right eye using a sampling rate of 1000Hz. Saccade sensitivity was set to high. Pupil size estimates indicate the pupil area. Gaze coordinates were recorded as screen x/y-positions. We implemented a drift check before each video started. If it failed (drift > 1° visual angle), the eye tracker was re-calibrated. Eye movements were only recorded while participants watched the educational videos.

2.5 Analysis

For the behavorial data, we did not exclude any participants or individual responses. If not otherwise noted, mean values and standard deviations are reported in the descriptive results.

For the gaze data, the raw Eyelink EDF files were loaded into EyeLink Data Viewer. A fixation report was generated for each participant across all four trials (videos) using the standard settings. In order to join the gaze data and pointer position data, we sampled both datasets at 1Hz (equals a bin size of 1 second) and calculated mean fixation or pointer position, respectively.

The pointer positions were calculated post-hoc as follows: In the postprocessing, the position of the laser pointer was extracted from the lecture videos by using Python 3.9.10 [Python.org 2022] with OpenCV 4.5.5.62 [OpenCV 2022], a library for computer vision. First, for each video frame, the laser pointer was segmented from the background. This was achieved by applying a color mask onto the video frame to subtract the background. The color mask maps on the range of red color shades representing the laser pointer. This results in an image including pixels in the color of this range of red color shades represented in white surrounded by black pixels. Then, OpenCV's blob detector [cv:SimpleBlobDetector 2022] was used to detect the laser pointer's position within the video frame. Danker et al. [Danker and Rosenfeld 1981] define a blob as a selective region with a luminance difference to the background. OpenCV's blob detector uses image thresholding to detect and group related pixels to form such regions in an image. Further, the algorithm provides parameters to filter for regions based on their color, size, and shape. In the case of laser pointer detection, these parameters were used to describe the laser pointer's size and shape. The algorithm returned the x and y position within the video frame of a detected matching blob, which was written together with the corresponding time in seconds of the video frame as comma-separated values to a file (CSV). When no matching blob was detected in the video frame, the x and y position was set to zero before being written to the CSV.

3 RESULTS

3.1 Quiz results

Overall, participants answered 78.53 9.26% of the quiz questions correctly. Correctness was highest for the hurricane topic (100 0%), followed by the nuclear plant topic (90.74 29.12%), electrolysis (74.17 43.96%) and nitrogen (48.81 50.29%). Since in the hurricane video condition, we cannot differentiate based on correctness, we will exclude this video from all further analyses involving correctness. Across all remaining topics, it made no difference to the correctness whether a presenter video feed was present (74.86 14.54%) or absent (72.72 16.48%), $t(22) = -0.607$, $p = .550$.

3.2 Fixations

We compared the average number and duration of fixations that fell on the presenter video feed or the content part of the slide (Area of Interest, AOI) separately for the two visibility conditions (Table 1). A repeated-measures ANOVA on the dependent variable fixation duration revealed the significant main effects of presenter visibility ($F(1,22) = 124.80$, $p < .001$, $\eta_p^2 = 0.850$), fixation AOI ($F(1,22) = 20.46$, $p < .001$, $\eta_p^2 = 0.482$) and their interaction ($F(1,22) = 157.45$, $p < .001$, $\eta_p^2 = 0.877$). A repeated-measures ANOVA on the dependent variable fixation number revealed a significant main effect of fixation AOI ($F(1,22) = 979.52$, $p < .001$, $\eta_p^2 = 0.978$) and the presenter visibility x AOI interaction ($F(1,22) = 9.61$, $p = .005$, $\eta_p^2 = 0.304$), but

Table 1: Mean (standard deviation) number of fixations per video per participant and mean (standard deviation) fixation duration per fixation noted separately for presenter visible and invisible conditions and whether the fixation where directed towards the content area or presenter area.

Area of Interest	Presenter visible		Presenter invisible	
	content	presenter	content	presenter
Fixation Number	599 (148)	80 (47)	670 (150)	12 (10)
Fixation Duration (ms)	305 (57)	482 (118)	322 (57)	237 (66)

not presenter visibility ($F(1,22) = 0.00$, $p < .965$, $\eta_p^2 < 0.001$). Further, there are no significant correlations between the correctness and either the total number of fixations on the presenter or average number of fixations on the presenter or the total fixation duration on the presenter or average fixation duration on the presenter (all *p*s > .1).

3.3 Gaze-to-Pointer Distance

We first visually inspected how the gaze-to-pointer distance was distributed over the course of time and see that it decreases over time (Figure A2), for all topics (see Figure A3), suggesting an orientation phase in the beginning.

In order to join gaze and pointer data, we normalized the screen coordinates to range from (0,0) in the top left to (1,1) in the bottom right, sampled both datasets at 1 Hz and calculated the average fixation position and pointer position for these 1-second-bins. One second gives the participant enough time to 'react' to changing pointer positions and adjust their gaze. For each of these bins, we calculated the x- and y- distances between the gaze position and pointer position as well as the point distance on the x/y-plane. In order to test our hypothesis that fixating the important parts of the slide (as indicated by the pointer) is important for quiz performance, we then aggregated both datasets to yield mean values for the normalized distances (x, y and point), their standard deviations and the percentage of correctly answered questions. This was done for each participant and for each except the hurricane video.

We then tested for correlations between the distance measures and correctness but did not find significant correlations (all ps > .38). We then calculated a multiple linear regression (Table 2) with the predictors *mean of the distance* and *standard deviation of the distance* and the outcome correctness (Model 1) and found that they explain a significant amount of the variance in the correctness ($F(2, 66) = 11.69$, $p < .001$, $R^2_{Adjusted} = .24$), with both the mean distance and the standard deviation of the distance being significant predictors. However, it is plausible to assume that the presenter visibility influences the mean distance to the pointer but not necessarily the correctness. Therefore, we calculated another regression model with the predictors *presenter visibility* (base: no presenter), *mean of the distance*, *standard deviation of the distance* and the two interactions of the visibility and distance measures (Model 2). The outcome variable was *correctness* again. We found that Model 2 explains a larger amount of the variance in the correctness ($F(5, 63) = 8.954$, $p < .001$, $R^2_{Adjusted} = .37$). To better interpret the interaction in Model 2, we plotted the data and regression lines for Model 2 separately for each presenter visibility condition (see Figure 3). Note, that we also ran the regression analyses only for those fixations directed at the content area. In this case, the mean distance is still predictive of the correctness and there is no predictive effect of an on-screen instructor anymore (see supplementary Table A3, Figure A3). Lastly, to explore future directions, we plotted the mean distance over time (see appendix, for all topics: Figure A2, separately: Figure A3), which is similarly going down over time for all topics.

4 DISCUSSION

In the present study, we investigated whether participants follow a digital laser pointer signal with their gaze and whether this pattern depends on the presence of an on-screen instructor. We found that the average distance between the learner's gaze and the pointer position can aid in predicting the student's quiz performance regardless of the presence of an on-screen instructor.

We explored whether looking at the instructor more often or longer has an impact on performance. For this, we compared the average fixation number and duration for the presenter area (in the upper right corner) and the content area. We found a significant interaction of area and instructor visibility: Fixations were directed less often at the content area when an on-screen instructor was visible (599) compared to when the instructor was not visible (670). The fixations directed at the presenter area were also held for longer when an instructor was visible (482 ms) compared to when no instructor was visible (237 ms). This is in line with previous studies showing an attention-capture of instructor video feeds [Stull et al. 2018; Wang et al. 2020]. Note, that the fixation duration on an on-screen instructor was by far the longest average fixation duration when comparing all conditions, which is not surprising given that it was the only 'moving' part of the videos.

We investigated potential effects of looking at an on-screen instructor on the post-test performance but we could not find any significant correlations. This seems to suggest that in our study, while the on-screen instructor had effects on the gaze patterns in general, this does not seem to have translated into a sizable detrimental effect with regards to post-test performance. In our regression model, instructor presence is negatively predicting the post-test performance, but only with an estimate of -1.53 %. According to a recent systematic review [Henderson and Schroeder 2021], that only reported positive effects or null results, our findings seem to give the very first indication that on-screen instructors might in some circumstances be detrimental to performance. However, the effect is so small we would consider it practically irrelevant. Additionally, the predictor is not significant, if we only look at the fixations directed towards the content area. A structured investigation into how detrimental effects of an on-screen instructor might depend on inter-individually mediated is certainly warranted.

Table 2: Estimates, confidence intervals and test statistics (with p-values denoted by stars) of the two linear models for predicting the outcome correctness. Model 1 employs the predictors mean of the distance and standard deviation of the distance, while Model 2 additionally employs the presenter visibility (base: presenter not visible) and its interaction with the two distance measures as predictors.

	Model 1			Model 2		
Predictors	*Estimates*	*95% CI*	*Statistic*	*Estimates*	*95% CI*	*Statistic*
(Intercept)	1.02 ***	0.65 – 1.38	5.57	1.99 ***	1.39 – 2.59	6.65
Distance Mean	-3.34 ***	-4.74 – -1.95	-4.80	-3.76 ***	-5.57 – -1.96	-4.16
Distance SD	3.09 **	1.21 – 4.96	3.29	-1.22	-4.82 – 2.37	-0.68
Presenter [visible]				-1.53 ***	-2.29 – -0.78	-4.07
Distance Mean: Presenter [visible]				1.45	-1.12 – 4.03	1.13
Distance SD: Presenter [visible]				5.43 *	1.21 – 9.66	2.57
R^2 / R^2 adjusted	0.262 / 0.239			0.415 / 0.369		

** p<0.05 ** p<0.01 *** p<0.001*

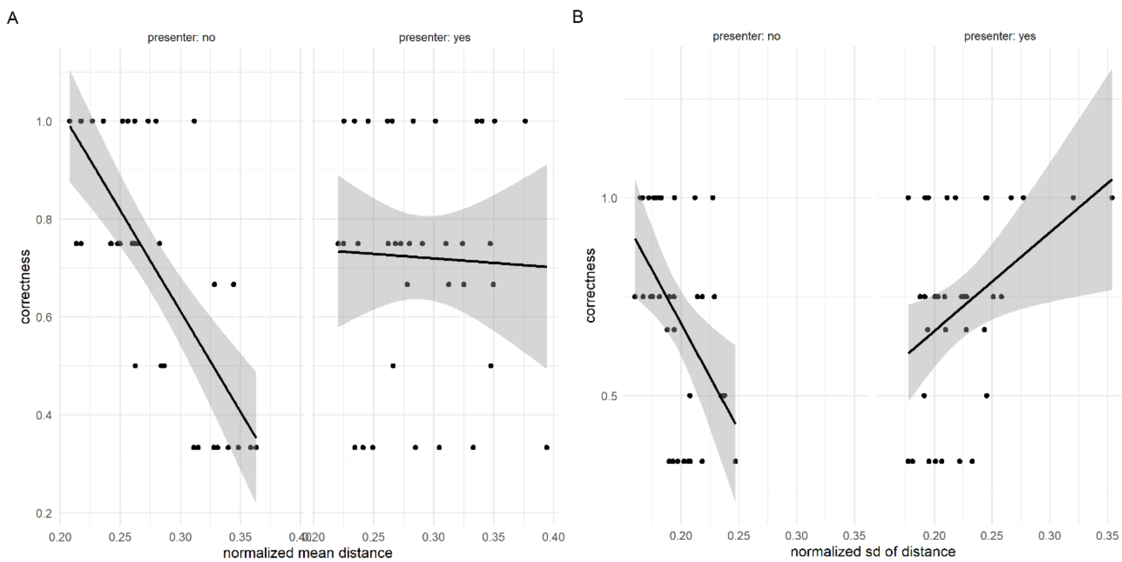

Figure 3: The figure shows the correctness (panel A: mean, panel B: standard deviation) for each participant and each video (except hurricane video) as a function of the average distance between gaze position and pointer separately for two presenter visibility conditions. Additionally, the regression line (black) is visualized together with the 95% confidence interval around the estimates (gray).

We then investigated the relationship between the gaze-to-pointer distance and the correctness. As separately calculated correlations did not yield significant results, in two steps, we calculated linear regression models. In Model 1 we used the average gaze-to-pointer distance and its standard deviation as predictors, in Model 2 we added the on-screen instructor visibility as a predictor. In both models, the mean gaze-to-pointer distance is a reliable predictor of correctness. Crucially, the predictive quality does not seem to be qualitatively dependent on the presence of an on-screen instructor. With each percent point of added gaze-to-pointer distance, correctness decreased by more than 3%. This is a practically sizable effect. It can aid in creating automated feedback systems, for example popping up a warning if, when watching a MOOC lecture video, a learner is consistently looking 15% away from the pointer (predicting a 50% worse post-test performance). This might happen, if the learner is inattentive in general. Another predictor for selective inattentiveness was thought to be the standard deviation of the gaze-to-pointer distance. It seems to be a predictor of post-test performance in Model 1 but with the added presenter visibility in Model 2, the standard deviation is only predictive in interaction with the presenter visibility. As can be seen in Figure 3, the standard deviation shows an opposite predictive pattern depending on presenter visibility. This seems to suggest that increased gaze-to-pointer distance is positive, but only when a presenter is visible and could potentially be due to positive affective effects an on-screen instructor has on the learner [Henderson and Schroeder 2021].

Notably, in the present explorative study, we used four different learning videos that were similar in style (all presented as pictures displaying a structured process of some sort). It is unclear how the results would translate to more lively educational videos with video content, as competition for attention is typically higher in dynamic settings which include many abruptly appearing objects [Franconeri and Simons 2003]. Additionally, we did not investigate a potential instructor model bias as on average, especially males, pay more attention to attractive faces and faces of desired gender (e.g. [Alexander and Charles 2009; Kranz and Ishai 2006]).

Overall, the present explorative study showed that gaze-to-pointer distance can be used as a general predictor of post-test performance in short educational videos (as they are often used in MOOCs), independent of an on-screen instructor. This may be beneficial in creating automated feedback-systems for learners using MOOCs and alert them of high gaze-to-pointer distance after finishing a video and suggest re-studying.

ACKNOWLEDGMENTS

We would like to thank Hannah Fleig for aiding in creating the stimulus material and Johannes Träg for helping with the data collection. This project is funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation, 425867974, to Anke Huckauf and Enrico Rukzio) and is part of Priority Program SPP2199 Scalable Interaction Paradigms for Pervasive Computing Environments.

REFERENCES

Gerianne M. Alexander and Nora Charles. 2009. Sex differences in adults' relative visual interest in female and male faces, toys, and play styles. *Archives of sexual behavior* 38, 3, 434–441. DOI: https://doi.org/10.1007/s10508-008-9429-7.

Märit Badman, Katja Höglund, and Odd V. Höglund. 2016. Student Perceptions of the Use of a Laser Pointer for Intra-Operative Guidance in Feline Castration. *Journal of veterinary medical education* 43, 2, 222–224.

Markus Bindemann, A. M. Burton, Ignace T. C. Hooge, Rob Jenkins, and Edward H. F. de Haan. 2005. Faces retain attention. *Psychon Bull Rev* 12, 6, 1048–1053. DOI: https://doi.org/10.3758/bf03206442.

Alan J. Danker and Azriel Rosenfeld. 1981. Blob detection by relaxation. *IEEE transactions on pattern analysis and machine intelligence* 3, 1, 79–92. DOI: https://doi.org/10.1109/tpami.1981.4767053.

SR Research - Fast, Accurate, Reliable Eye Tracking. 2021. *EyeLink 1000 Plus* (August 2021). Retrieved January 28, 2022 from https://www.sr-research.com/eyelink-1000-plus/.

Steven L. Franconeri and Daniel J. Simons. 2003. Moving and looming stimuli capture attention. *Perception & Psychophysics* 65, 7, 999–1010. DOI: https://doi.org/10.3758/bf03194829.

Marsha L. Henderson and Noah L. Schroeder. 2021. A Systematic review of instructor presence in instructional videos: Effects on learning and affect. *Computers and Education Open* 2, 100059. DOI: https://doi.org/10.1016/j.caeo.2021.100059.

Jianzhong Hong, Zhongling Pi, and Jiumin Yang. 2018. Learning declarative and procedural knowledge via video lectures: cognitive load and learning effectiveness. *Innovations in Education and Teaching International* 55, 1, 74–81. DOI: https://doi.org/10.1080/14703297.2016.1237371.

Hubert Knoblauch. 2008. The performance of knowledge: Pointing and knowledge in Powerpoint presentations. *Cultural sociology* 2, 1, 75–97.

Felicitas Kranz and Alumit Ishai. 2006. Face perception is modulated by sexual preference. *Current biology : CB* 16, 1, 63–68. DOI: https://doi.org/10.1016/j.cub.2005.10.070.

Stephen R. H. Langton, Anna S. Law, A. M. Burton, and Stefan R. Schweinberger. 2008. Attention capture by faces. *Cognition* 107, 1, 330–342. DOI: https://doi.org/10.1016/j.cognition.2007.07.012.

Jens Madsen, Sara U. Júlio, Pawel J. Gucik, Richard Steinberg, and Lucas C. Parra. 2021. Synchronized eye movements predict test scores in online video education. *Proceedings of the National Academy of Sciences of the United States of America* 118, 5. DOI: https://doi.org/10.1073/pnas.2016980118.

Sebastiaan Mathôt, Daniel Schreij, and Jan Theeuwes. 2012. OpenSesame: an open-source, graphical experiment builder for the social sciences. *Behav Res* 44, 2, 314–324. DOI: https://doi.org/10.3758/s13428-011-0168-7.

OpenCV. 2022. *Home - OpenCV* (January 2022). Retrieved January 28, 2022 from https://opencv.org/.

cv:SimpleBlobDetector *Class Reference* (January 2022). Retrieved January 28, 2022 from https://docs.opencv.org/4.5.5/d0/d7a/classcv_1_1SimpleBlobDetector.html.

OpenSesame *// OpenSesame documentation* (January 2022). Retrieved January 28, 2022 from https://osdoc.cogsci.nl/.

Zhongling Pi, Yi Zhang, Fangfang Zhu, Ke Xu, Jiumin Yang, and Weiping Hu. 2019. Instructors' pointing gestures improve learning regardless of their use of directed gaze in video lectures. *Computers & Education* 128, 345–352. DOI: https://doi.org/10.1016/j.compedu.2018.10.006.

Python.org. 2022. *Welcome to Python.org* (January 2022). Retrieved January 28, 2022 from https://www.python.org/.

Kshitij Sharma. 2015. *Gaze analysis methods for learning analytics*. EPFL.

Oleg Špakov, Diederick Niehorster, Howell Istance, Kari-Jouko Räihä, and Harri Siirtola. 2019. Two-Way Gaze Sharing in Remote Teaching. In . Springer, Cham, 242–251. DOI: https://doi.org/10.1007/978-3-030-29384-0_16.

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, 263–272. DOI: https://doi.org/10.1016/j.chb.2018.07.019.

Eunmo Sung and Richard E. Mayer. 2012. Affective impact of navigational and signaling aids to e-learning. *Computers in Human Behavior* 28, 2, 473–483. DOI: https://doi.org/10.1016/j.chb.2011.10.019.

Margot van Wermeskerken, Susanna Ravensbergen, and Tamara van Gog. 2018. Effects of instructor presence in video modeling examples on attention and learning. *Computers in Human Behavior* 89, 430–438. DOI: https://doi.org/10.1016/j.chb.2017.11.038.

Jiahui Wang, Pavlo Antonenko, and Kara Dawson. 2020. Does visual attention to the instructor in online video affect learning and learner perceptions? An eye-tracking analysis. *Computers & Education* 146, 103779. DOI: https://doi.org/10.1016/j.compedu.2019.103779.

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

Zhonggen Yu. 2021. The effect of teacher presence in videos on intrinsic cognitive loads and academic achievements. *Innovations in Education and Teaching International*, 1–12. DOI: https://doi.org/10.1080/14703297.2021.1889394.

A APPENDIX

A.1 Additional results tables and figures

Table A3: Estimates, confidence intervals and test statistics (with p-values denoted by stars) of the two linear models for predicting the outcome correctness. Model 1 employs the predictors mean of the distance and standard deviation of the distance, while Model 2 additionally employs the presenter visibility (base: presenter not visible) and its interaction with the two distance measures as predictors.

	correctness			correctness		
Predictors	*Estimates*	*CI*	*Statistic*	*Estimates*	*CI*	*Statistic*
(Intercept)	1.80 ***	1.32 – 2.29	7.43	2.39 ***	1.61 – 3.17	6.10
Distance Mean	-3.15 ***	-4.90 – -1.39	-3.58	-3.32 *	-5.91 – -0.73	-2.56
Distance SD	-1.11	-5.04 – 2.81	-0.57	-4.05	-10.39 – 2.28	-1.28
Presenter [visible]				-0.92	-1.92 – 0.08	-1.83
Distance Mean: Presenter [visible]				0.69	-2.84 – 4.21	0.39
Distance SD: Presenter [visible]				3.99	-4.13 – 12.11	0.98
Observations	69			69		
R^2 / R^2 adjusted	0.376 / 0.357			0.417 / 0.371		
* *p<0.05* ** *p<0.01* *** *p<0.001*						

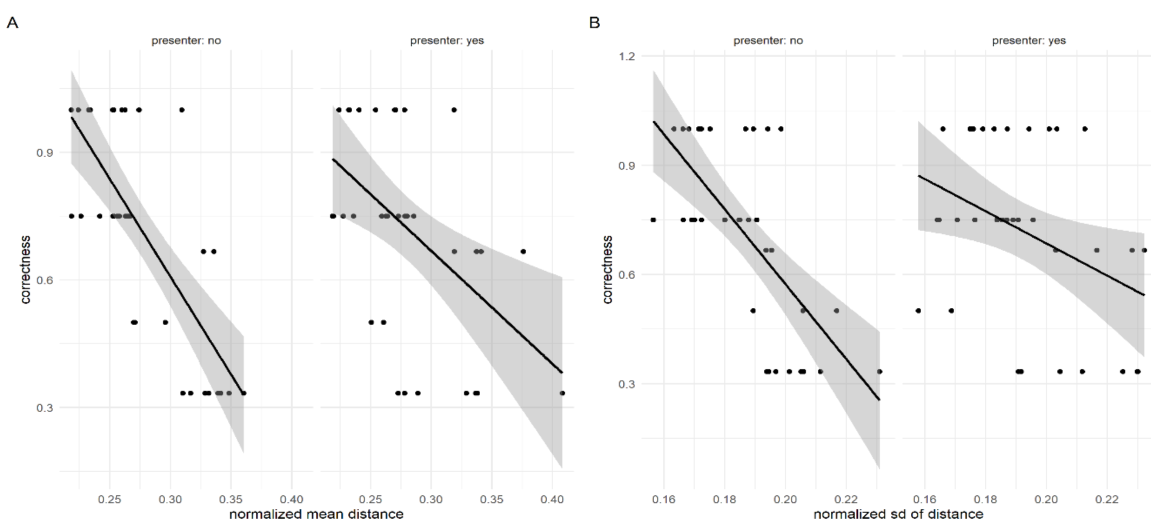

Figure A1: The figure shows the correctness (panel A: mean, panel B: standard deviation) for each participant and each video (except hurricane video) as a function of the average distance between gaze position and pointer separately for two presenter visibility conditions. Additionally, the regression line (black) is visualized together with the 95% confidence interval around the estimates (gray).

A.2 Text elements of the experiment in original language (German)

A.2.1 Willkommenstext. Liebe/r Teilnehmer/in,

herzlich Willkommen und vielen Dank für dein Interesse an dieser Studie. Die Studie hat das Ziel zu untersuchen, wie die Gestaltung von Online-Lernvideos optimiert werden kann. Durch deine Teilnahme leistest du einen wichtigen Beitrag zur Wissenschaft und unterstützt uns bei unserer Forschung.

Alle Daten, die in dieser Studie erhoben werden, werden von uns vertraulich behandelt und in pseudomisierter Form gespeichert, sodass keine Rückschlüsse auf deine Person möglich sind. Bitte sei also bei der Beantwortung der Fragen der Studie ehrlich.

Herzlichen Dank für deine Unterstützung und viel Spaßbei der Studie!

A.2.2 Demographische Fragen. Folgende demographische Fragen wurden zu Beginn der Studie gestellt.

Bitte gebe dein Alter als Zahl in Jahren an.

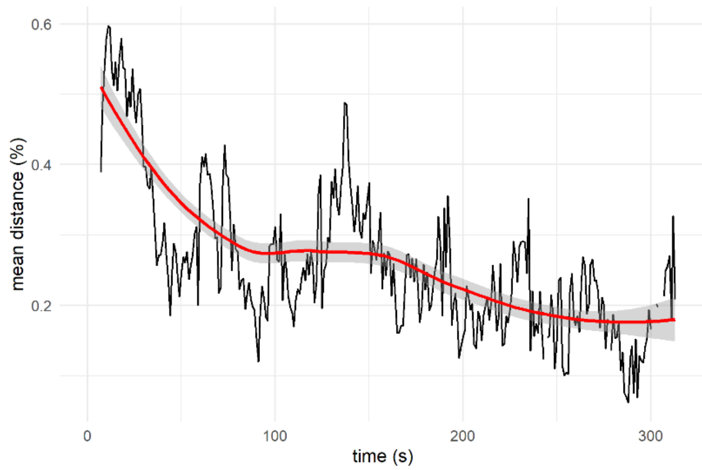

Figure A2: The figure depicts the average normalized gaze-to-pointer distance (averaged over all participants and topics) as a function of the time in seconds (black line). Gaps in the data line indicate missing pointer data (e.g. no pointer visible). The red curve indicates a loess-smoothed approximation (gray surround: 95% confidence interval).

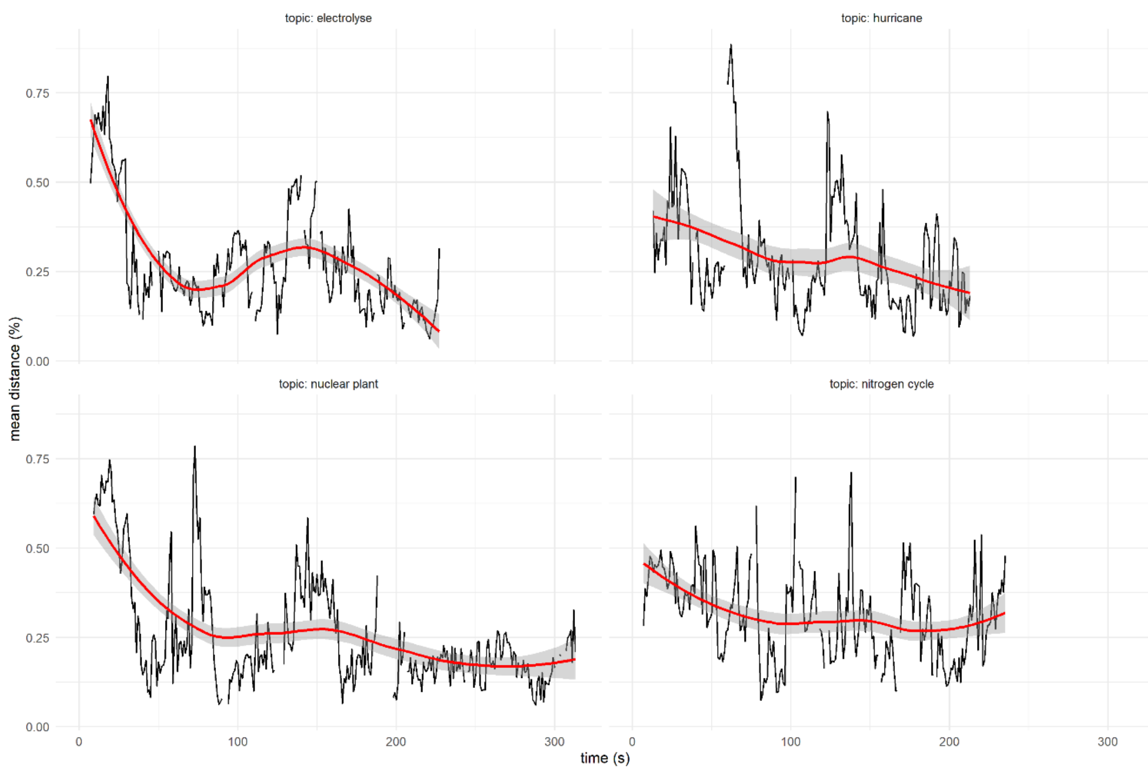

Figure A3: The figure depicts the average normalized gaze-to-pointer distance (averaged over all participants and separately for each topic) as a function of the time in seconds (black line). Gaps in the data line indicate missing pointer data (e.g. no pointer visible). The red curve indicates a loess-smoothed approximation (gray surround: 95% confidence interval).

Welchem Geschlecht fühlst du dich zugehörig?

männlich
weiblich
divers
keine Angabe

Wie gut beherrschst du die deutsche Sprache?

1. Deutsch ist meine Muttersprache.
Ich beherrsche sie nahezu auf dem Niveau eines/r Muttersprachlers/in.
Ich beherrsche sie verhandlungssicher.
Ich beherrsche sie fließend.
Ich beherrsche sie gut/sicher.
Ich besitze fortgeschrittene Grundkenntnisse.
Ich besitze Grundkenntnisse oder schlechter.

Bitte gebe deinen höchsten Bildungsabschluss an.

1. Kein Schulabschluss
Hauptschulabschluss
Realschulabschluss/ Mittlere Reife
Fachabitur/ Fachhochschulreife
Allgemeine Hochschulreife/ Abitur
Abgeschlossene Berufsausbildung
Fachhochschulabschluss
Hochschulabschluss/ Universitätsabschluss

Bist du derzeit als Student/in an einer Hochschule oder Universität eingeschrieben?

1. Ja
Nein

A.2.3 Instruktionstext vor dem Experiment. Nun beginnt das eigentliche Experiment.

Du wirst nun nacheinander insgesamt vier kurze Lernvideos ansehen. Jedes Video wird nur genau ein Mal gezeigt und es kann nicht pausiert werden. Nach jedem Video folgt zuerst eine kurze Frage zu deinem persönlichen Eindruck über das Video und anschließend eine Wissensabfrage über den Inhalt des vorherigen Videos. Bitte schaue dir alle Videos aufmerksam an und beantworte alle Wissensfragen so gut du kannst.

Während des Abspielens der Videos werden mit Hilfe des Eyetrackers deine Augenbewegungen aufgezeichnet.

A.2.4 Fragen nach jedem Video. Folgende Fragen wurden nach jedem der vier Videos gestellt.

Wie neu waren die Inhalte des eben gesehen Videos für dich?

1. Sehr neu
2. 2
3. 3
4. 4
5. Alles bekannt

Bitte gib an, wie schwierig du den Inhalt des Lernvideos empfandest.

1. Sehr leicht
Leicht
Eher leicht
Ausgewogen
Eher schwer
Schwer
Sehr schwer

Wie sehr hat dir die Darstellungsform des Videos zugesagt?

1. Die Darstellungsform hat mir sehr zugesagt.
Die Darstellungsform hat mir eher zugesagt.
Die Darstellungsform hat mir mittelmäßig zugesagt.
Die Darstellungsform hat eher nicht zugesagt.
Die Darstellungsform hat überhaupt nicht zugesagt.

A.2.5 Fragen zur Elektrolyse. Folgende Fragen wurden nach dem Video über die Elektrolyse gestellt.

Welche Aussage ist richtig? (Nur eine Antwort ist richtig)

1. An der Kathode nehmen Kationen Elektronen auf. An der Anode geben Anionen Elektronen an die Anode ab.
An der Anode nehmen Kationen Elektronen auf. An der Kathode geben Anionen Elektronen an die Kathode ab.
An der Kathode nehmen Anionen Elektronen auf. An der Anode geben Kationen Elektronen an die Anode ab.
An der Anode nehmen Anionen Elektronen auf. An der Kathode geben Kationen Elektronen an die Kathode ab.

Welche Aussage bezüglich der Elektrolyse von Zinkiodid ist richtig? (Eine Antwort ist richtig)

1. Die positiv geladenen Zinkionen sind Teil der Reduktion.
Die negativ geladenen Zinkionen sind Teil der Reduktion.
Die negativ geladenen Zinkionen sind Teil der Oxidation.
Die positiv geladenen Zinkionen sind Teil der Oxidation.

Bitte wähle die Aussagen aus, die wahr sind. (Mehrere Antworten können richtig sein)

1. An der Anode herrscht ein Elektronenüberschuss.
An der Anode herrscht ein Elektronenmangel.
An der Anode ist der Pluspol angelegt.
An der Anode ist der Minuspol angelegt.

Bitte wähle die Aussagen aus, die wahr sind. (Mehrere Antworten können richtig sein)

1. Bereits bei einer geringen angelegten Spannung erfolgen beide Prozesse der Elektrolyse.
Durch das Anlegen einer Wechselspannung können die Prozesse der Elektrolyse beschleunigt werden.
Elektroden bestehen immer aus Metall.
Kationen sind positiv geladene Ionen.

A.2.6 Fragen zu Hurrikans. Folgende Fragen wurden nach dem Video über Hurrikans gestellt.

Welche Aussage ist richtig? (Nur eine Antwort ist richtig)

1. Zur Entstehung von Hurrikans muss die Lufttemperatur etwa 26,5°C betragen.
Zur Entstehung von Hurrikans muss die Wassertemperatur etwa 26,5°C betragen.
Zur Entstehung von Hurrikans muss die Lufttemperatur etwa 21,5°C betragen.
Zur Entstehung von Hurrikans muss die Wassertemperatur etwa 21.5°C betragen.

Tropische Wirbelstürme können unterschiedlich genannt werden: Hurricanes, Taifuns oder Zyklonen. Wovon sind die unterschiedlichen Bezeichnungen abhängig? (Eine Antwort ist richtig)

1. Von der Himmelsrichtung in die sich der Wirbelsturm bewegt.

Von der maximalen Geschwindigkeit, die erreicht werden kann.
Von dem Entstehungsort.
Von der Entfernung vom Äquator.

Was passiert bei der Entstehung von Hurrikans? (Eine Antwort ist richtig)

1. Kalte Luft sinkt nach unten.
   Kalte Luft steigt nach oben.
   Warme Luft sinkt nach unten.
   Warme Luft steigt nach oben.

Weswegen bewegen sich Hurrikans fort? (Eine Antwort ist richtig)

1. Durch die Corioliskraft.
   Durch die Zentrifugalkraft.
   Durch Meeresströmungen.
   Durch Passatwinde.

A.2.7 Fragen zum Kernkra werk. Folgende Fragen wurden nach dem Video über Kernkraftwerke gestellt.

Kernkraftwerke haben drei verschiedene Wasserkreisläufe. Die Namen dieser lauten: (Nur eine Antwort ist richtig)

1. Primärkreislauf, Sekundärkreislauf, Tertiärkreislauf.
   Reaktorkreislauf, Maschinenhauskreislauf, Kühlkreislauf.
   Primärkreislauf, Sekundärkreislauf, Kühlkreislauf.
   Reaktorkreislauf, Sekundärkreislauf und Kühlkreislauf.

Die Atomkerne werden gespalten durch... (Eine Antwort ist richtig)

1. ... Elektronen.
   ... Neutronen.
   ... Protonen.
   ... hohen Druck.

Gib an, welche der folgenden Aussagen richtig ist. (Eine Antwort ist richtig)

1. Aus der Kernenergie entsteht erst Wärmeenergie, dann Bewegungsenergie und zuletzt elektrische Energie.
   Aus Kernenergie entsteht erst Bewegungsenergie, dann Wärmeenergie und zuletzt elektrische Energie.
   Aus der Kernenergie entsteht erst Wärmeenergie, dann elektrische Energie und zuletzt Bewegungsenergie.
   Aus der Kernenergie entsteht erst Bewegungsenergie, dann elektrische Energie und zuletzt Wärmeenergie.

Gib an, welche der folgenden Aussagen richtig sind. Kernkraftwerke erzeugen... (Mehrere Antworten können richtig sein)

1. ... große Menge an Abwärme, die an die Umwelt abgegeben werden müssen.
   ... giftige Abgase, die durch den Kühlturm abgegeben werden müssen.
   ... elektrischen Strom.
   ... schweres Wasser.

A.2.8 Fragen zum Sticksto kreislauf. Folgende Fragen wurden nach dem Video über den Stickstoffkreislauf gestellt.

Wie setzt sich die Luft zusammen? (Nur eine Antwort ist richtig)

1. 68% Stickstoff, 31% Sauerstoff, 1% Edelgase und Kohlenstoffdioxid
   68% Stickstoff, 21% Sauerstoff, 11% Edelgase und Kohlenstoffdioxid
   78% Stickstoff, 11 % Sauerstoff, 11% Edelgase und Kohlenstoffdioxid
   78% Stickstoff, 21% Sauerstoff, 1% Edelgase und Kohlenstoffdioxid

Welche der untenstehenden Begriffe sind zentrale Schritte des Stickstoffkreislaufes? (Mehrere Antworten können richtig sein)

1. Ammonifikation
   Nitrifikation
   Entnitrifikation
   Stickstofffixierung

Gib an, welche der folgenden Aussagen richtig sind. (Mehrere Antworten können richtig sein)

1. Bei der Stickstofffixierung entsteht durch die Symbiose von Stickstofffixierenden Bakterien und Pflanzenwurzeln Ammonium.
   Pflanzen können Ammonium für Aminosäuren, Proteine oder andere Verbindungen aufnehmen.
   Tiere nehmen Stickstoff über das Essen von Pflanzen auf.
   Das Ausscheiden des Stickstoffes von Tieren sorgt dafür, dass der Schritt der Stickstofffixierung komplett ausgelassen wird.

A.2.9 Abschlussfragen. Folgende Fragen wurden abschließend gestellt.

Vergleiche nun alle Videos noch einmal hinsichtlich der Schwierigkeit ihrer Inhalte, indem du eine Rangreihe bildest. Bitte gebe dem Video, welches du als am leichtesten empfandest die Nummer 1, und dem Video welches du als am schwersten empfandest die Nummer 4. Den anderen Videos teilst du die Nummern 2 und 3 zu. Bitte vergebe jede Nummer (1, 2 ,3 und 4) genau ein Mal. Versuche dich auch hier wieder nur auf die Lernvideos zu beziehen und nicht darauf, ob du das Beantworten der anschließenden Wissensfragen als schwierig oder leicht empfunden hast. Es kommt hierbei nicht darauf an, dass du dich perfekt an alle Videos zurückerinnerst.

Wie schwer ist es dir gefallen die Rangreihe zu bilden?

1. sehr schwer
   eher schwer
   durchschnittlich
   eher leicht
   sehr leicht

Schätze nun, wieviel der Zeit du in Richtung Laserpointer oder Sprecher oder irgendwo sonst hingeschaut hast.

Von der Zeit, die du entweder auf Sprecher oder Laserpointer geschaut hast, schätze nun, wieviel der Zeit du in Richtung Laserpointer oder Sprecher hingeschaut hast.

Wie war sonst dein Eindruck von der Einblendung des Präsentators oder der Benutzung des Laserpointers?