Scalable Mind-Wandering Detection for MOOCs: A Webcam-Based Approach

Yue Zhao, Christoph Lofi, and Claudia Hauff*

Delft University of Technology, Web Information Systems, Delft, the Netherlands
{y.zhao-1,c.lofi,c.hauff}@tudelft.nl

Abstract. Mind-wandering or loss of focus is a frequently occurring experience for many learners and negatively impacts learning outcomes. While in a classroom setting, a skilled teacher may be able to react to students' loss of focus, in Massive Open Online Courses (MOOCs) no such intervention is possible (yet). Previous studies suggest a strong relationship between learners' mind-wandering and their gaze, making it possible to detect mind-wandering in real-time using eye-tracking devices. Existing research in this area though has made use of *specialized* (and expensive) hardware, and thus cannot be employed in MOOC scenarios due to the inability to scale beyond lab settings. In order to make a step towards *scalable* mind-wandering detection among online learners, we propose the use of ubiquitously available consumer grade webcams. In a controlled study, we compare the accuracy of mind-wandering detection from gaze data recorded through a standard webcam and recorded through a specialized and high-quality eye tracker. Our results suggest that a large-scale application of webcam-based mind-wandering detection in MOOCs is indeed possible.

Keywords: learning analytics, MOOCs, mind-wandering, eye tracking

1 Introduction

Mind-wandering is an essential part of human behavior consuming up to 50% of everyday thoughts [8], and can be described as "thoughts and images that arise when attention drifts away from external tasks and perceptual input toward a more private, internal stream of consciousness" [12]. While mind-wandering can also have positive effects (such as fostering creativity [23]), many educational tasks including following a lecture or solving an assignment require active attention and focus to reach the desired learning outcomes. For these tasks, excessive mind-wandering has disastrous effects on learning efficiency [19].

In the traditional classroom setting, mind-wandering and attention lapses have been studied for a long time, e.g. [3, 24]. Although researchers do not yet

* This work is partially supported by the *Leiden-Delft-Erasmus Centre for Education and Learning.*

agree on the actual attention span of learners, several past works have found attention among students during lecture time to vary in a cyclic manner.

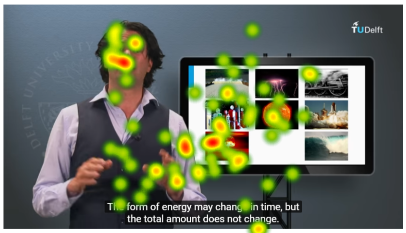
For online courses and MOOCs, this problem is even more severe as they are consumed using digital display devices. This mode of consumption is particularly prone to mind-wandering. Likely due to the ubiquity of smartphones and digital content, a significant subgroup of online users adopt a "heavy media multitasking" behavior [10], making it challenging for them to focus on a single multimedia content unit. This finding is also supported by our work, where learners frequently lose focus even in short video clips of around seven minutes.

In order to detect mind-wandering among online learners during their consumption of digital materials, we require an approach that is *scalable* (it can be deployed to thousands of learners), *near real-time* (mind-wandering is detected as soon as it occurs), *unobtrusive* (learners are not distracted by the detection procedure) and *autonomous*. In addition to providing insights into learners' behaviors, such a method would also enable real-time interventions that lower the amount of mind-wandering taking place. As a concrete example we envision an intelligent MOOC video player: the player (via the webcam feed) monitors a learner's attention state and when a loss of focus is detected, the player pauses the video automatically in order to avoid skipping over relevant content. In order to ensure learners' privacy, all necessary processing will be client-side (i.e. executed within the browser).

To this end, previous research showed that by analyzing people's gaze data, mind-wandering can be detected, e.g. whilst reading texts on screen [1], or watching (non-educational) films [2]. These results can be attributed to the eye-mind link effect [15], which states that "there is no appreciable lag between what is fixated and what is processed." Existing works usually rely on expensive and specialized eye-tracking hardware (e.g. a Tobii eye tracker) to obtain gaze data, which is not available to the average MOOC learner. It is therefore still an open question whether eye-tracking based mind-wandering detection can be performed in a scalable manner.

Our goal in this paper is to develop a fully automatic method for detecting mind-wandering and loss of focus in near real-time using only low-end webcams ubiquitously found on laptop computers. To this end, we conducted a laboratory study with 13 participants, collecting a dataset of gaze features (i.e. features extracted from gaze data) and self-reported mind-wandering. To motivate this approach, refer to Fig. 1 which visualizes the gaze of two of our study participants through heatmaps. The MOOC video shown has several relevant visual areas, including the lecture slides, the subtitles, and the speaker's face. In the depicted scene, a changing set of examples is shown on the slides which are important to grasp the lecture content. The participant who reported mind-wandering in the 30s interval intently gazed on a spot on the speaker's face, ignoring the slides and the shown examples, while the second participant who reported no mind-wandering focused on all relevant areas of the video. Our proposed approach employs supervised machine learning to automatically learn such mind-wandering patterns based on gaze features.

(a) reported mind-wandering



(b) no mind-wandering

Fig. 1: Gaze heatmaps of two study participants over a 30 second interval

Our contributions in this work are as follows:

1. We create an elaborate gold dataset to foster eye-tracking based mind-wandering research, featuring 13 participants watching two MOOC videos each in a controlled laboratory setting, reporting feedback on mind-wandering in brief intervals. In addition to these mind-wandering reports, we provide video and gaze data as recorded and analyzed by a professional eye tracker as well as gaze data recorded by a webcam and processed by an open-source gaze library. We make this data available on our companion Web page [25].
2. We implement and evaluate an approach to automatically detect mind-wandering based on gaze data (i) collected with a specialized eye-tracking device (`Tobii X2-30`), relying on the results and best practices published in [2], and (ii) collected with a standard webcam.
3. We extensively discuss and evaluate both approaches, and argue that our webcam-based method is indeed suitable for large-scale deployment outside a controlled laboratory setting.

2 Background: Mind-Wandering

Different data collection methods have been used to study mind-wandering of students in traditional classrooms since the 1960s, such as the observation of inattention behaviors [7], the retention of course content [11], using direct probes in class [9, 21] and relying on self-reports from students [3]. A common belief was that learners' attention may decrease considerably after 10-15 minutes of the lecture, which was supported by [21]. However, Wilson and Korn [24] later challenged this claim and argued that more research is needed. In a recent study, Bunce et al. [3] asked learners to report their mind-wandering voluntarily during 9-12 minute course segments. Three buttons were placed in front of each learner, representing attention lapses of 1 minute or less, of 2-3 minutes and of 5 minutes or more. During the lectures, the learners were asked to report their mind-wandering by pressing one of three buttons once they *noticed* their mind-wandering. This setup led Bunce et al. [3] to conclude that learners start losing their attention early on in the lecture and may cycle through several attention states within the 9-12 minute course segments.

In online learning environments, mind-wandering may be even more frequent. Risko et al. [16] used three one hour video-recorded lectures with different topics (psychology, economics, and classics) in their experiments. While watching the videos, participants were probed four times throughout each video. The mind-wandering frequency among the participants was found to be 43%. Additionally, Risko et al. [16] found a significant negative correlation between test performance and mind-wandering. Szpunar et al. [22] investigated the impact of interpolated tests on learners' mind-wandering within online lectures. The study participants were asked to watch a 21-minute video lecture (4 segments with 5.5 minutes per segment) and report their mind-wandering in response to random probes (one probe per segment). In their experiments, the mind-wandering frequency was about 40%. Loh et al. [10] also employed mind-wandering probes to measure learners' mind-wandering and found a positive correlation between media multitasking activity and learners' mind-wandering (average frequency of 32%) whilst watching video lectures. Based on these considerably high mind-wandering frequencies we conclude that reducing mind-wandering in online learning is an important approach to improve learning outcomes.

Inspired by the eye-mind link effect [15], a number of previous studies [1,2,13] focused on the automatic detection of learners' mind-wandering by means of gaze data. In [1, 2], Bixler and D'Mello investigated the detection of learners' mind-wandering during computerized reading. To generate the ground truth, the study participants were asked to manually report their mind-wandering when an auditory probe (i.e. a beep) was triggered. Based on those reports, the mind-wandering frequency ranged from 24.3% to 30.1%. During the experiment, gaze data was collected using a dedicated eye tracker. In [13], Mills et al. asked the study participants to watch a 32 minute, non-educational movie and self-report their mind-wandering throughout. In order to detect mind-wandering automatically, statistical features and the relationship between gaze and video content were considered. In contrast to [1,2], the authors mainly focused on the relationship between a participant's gaze and areas of interest (AOIs), specific areas in the video a participant should be interested (like the speaker or slides).

3 Methodology

In our study, we focus on the automatic detection of learners' mind-wandering through webcam-based eye tracking. The scenario we consider is video lecture watching, which is the most common manner of conveying lecture content in MOOCs [16]. We collect data through a lab study with 13 participants who were asked to watch two lecture videos and regularly report their mind-wandering during this time. We recorded their gaze data with a dedicated high-quality eye tracker and a standard webcam. In our paper, gaze data refers to both gaze points (the points on the screen a participant is actively looking at) and gaze events (i.e. fixations and saccades). Fixation refers to the action that concentrates the gaze points on a single area, and saccade refers to the quick and simultaneous movement of both eyes between two or more phases of fixations.

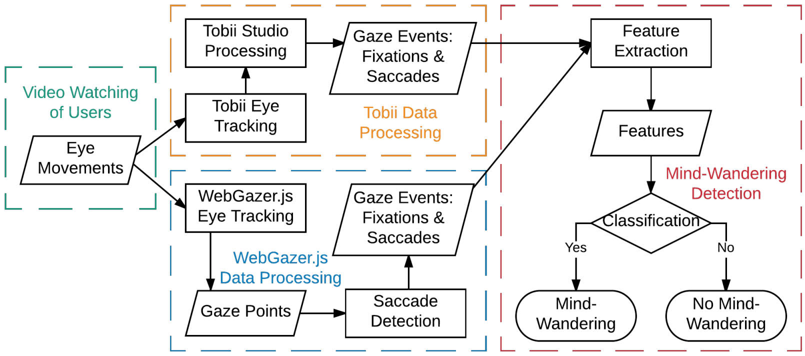


Fig. 2: Overview of the processing pipeline

Compared to previous works [10, 13, 16, 22], the two MOOC lecture videos in our study are considerably shorter - they are between six and eight minutes in length, in line with standard MOOC practices today. To collect the ground-truth (did mind-wandering occur in the last n seconds?) we rely on mind-wandering probes which have proven to be effective in the traditional classroom setting [4, 9, 21] and online learning [1, 2]. Probes (regularly and actively seeking input from the study participants) are more reliable than self-caught reports which require study participants to think about their loss of focus and about reporting it [20]. In response to our probes (in the form of an auditory signal — a bell) during video lecture playback, participants were asked to press a key to indicate that they experienced mind-wandering in the past 30 seconds. Participants who did not experience mind-wandering were asked to ignore the bell and continue watching.

Having collected the ground truth data, we next turned to the extraction of features from gaze data, following [13]. In line with previous works, we extracted features from gaze events. These gaze events are generated by gaze points. Note that gaze points are not measured directly - they are estimated from the recorded eye and iris movements; we used the existing software libraries of our dedicated high-quality eye tracker and our open-source webcam-based framework to turn eye and iris movements into gaze points.

Finally we used employed the ground truth data and extracted features in a supervised machine learning task to explore to what extent the automatic detection of mind-wandering in this setting is possible.

The overview of the processing pipeline is shown in Fig. 2. In the following sections, we first describe in more detail the experimental design of our study, and then elaborate on the features we extracted.

3.1 Study Setup

Our study is built around two introductory videos taken from two different x-MOOCs [17] professionally produced and offered by the Delft University of

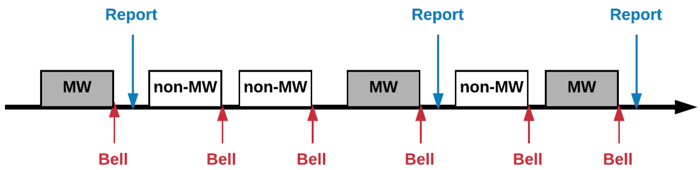


Fig. 3: An example mind-wandering report

Technology on the edX[1] platform. One video, (taken from the *Understanding Nuclear Energy* MOOC), covers the basics of the atomic model with a length of 6:41 minutes; the second one (part of the *Solar Energy* MOOC and 7:49 minutes long) introduces the concept of energy conversion. We selected those videos specifically as they contain rich visual lectures slides overlayed with the speaker (see Fig. 1). They cover topics we consider interesting to a wider audience and do not require extensive prior knowledge due to their introductory nature. All study participants watched both videos; their order was randomized to avoid order effects.

We used two eye-tracking devices in the study, a high-quality one as a reference and a low-quality webcam. Concretely, we made use of the professional `Tobii X2-30` eye tracker and its corresponding software Tobii Studio to estimate participants' gaze points. Our webcam is the built-in camera of our experimental laptop, a Dell Inspiron 5759 with a 17-inch screen and a 1920×1080 resolution. To estimate the gaze points based on a live webcam feed, we relied on `WebGazer.js` [14], an open source eye-tracking library written in JavaScript. We built a Web application closely resembling existing MOOC lecture video players with additional logging capabilities. In order to alert our participants to each mind-wandering probe, we included a medium-volume acoustic bell signal played by the Web application. After the bell, participants reported their mind-wandering in the past 30 seconds by pressing a feedback button. The next bell signal occurred after another 30-60 seconds. The actual time was randomized within those boundaries, as previous research [1, 10] suggests that participants perceive interruptions which are not perfectly periodic as less interrupting. In order to further limit the mental annoyance of this process, participants were only asked to actively report in case they had indeed experienced mind-wandering. This process resulted in mind-wandering reports for each participant, including the bell signals and participant responses with respect to mind-wandering as shown in Fig. 3.

We recruited our study participants (six females, seven males, all with a computer science background) through an internal mailing list and did not pay them. After a pre-study briefing, we asked our participants, six of whom wore glasses or contact lenses, to sit stable and comfortably in front of the laptop (with a distance of $52 - 68$ cm between eyes and screen). The study consisted of pre- and post-study questionnaires, an instruction phase by the experimenter,

[1] `http://edx.org/`

a calibration phase (to calibrate the eye trackers) and the watching of the two lecture videos; overall, participants spent about 35 minutes in the experiment. We conducted all experiments during daylight hours with both office lights and natural daylight contributing to our lighting.

The data generated by Tobii Studio during the study includes (among others) the estimated 2D coordinates of gaze points for each eye, the duration and coordinates of gaze events (i.e. fixations and saccades), the eye and pupil positions of the participant as well as the distance between the participant and the camera with a sample rate of 30 samples/second. In contrast, the data extracted from our webcam-based eye-tracking solution only includes the estimated 2D coordinates of gaze points of both eyes sampled at a rate of 5 samples/second.

3.2 Mind-wandering Detection using Gaze Features

To realize eye-tracking based mind-wandering detection using the professional eye tracker and our webcam-based solution, we turn the task into a standard supervised machine learning task. Our classifiers are trained using the aforementioned mind-wandering reports as reference labels, and extracted gaze features for each time span between two bell signals as collected by either technique as input.

Given Tobii Studio's gaze data and inspired by [1,2] we extracted 58 features in total. These features can be classified into two groups, global features and local features. The global features refer to features which are independent of the current content of the MOOC video, and are as shown in Tab. 1 based on fixations and saccades. The feature vector of a given bell time span covers statistical aggregates of fixation and saccade data such as maximum, minimum, mean, median, standard deviation, range, kurtosis and skew of fixation durations, saccade durations, saccade distance and saccade angles.

Local features are mainly based on the relationship between fixations/saccades and the areas of interest (AOIs) in the MOOC video, i.e. local features correlate gaze data with the current video content. There are certain areas of a video where a focused learner should focus her attention (e.g. the slides) in order to follow the content, while others are less interesting. While this opens a complex design space for engineering features, we opted for a simple implementation in which we manually defined three fixed areas of interest: the instructor's face, subtitles, and the lecture slides. The resulting local features include then the number and length of saccades and fixations which focus on different areas of interest for a given time span. Recall once more that all saccade and fixation data are computed by Tobii Studio with high precision for each bell time span based on a raw sample rate of 30 Hz.

Due to limitations of the `WebGazer.js` framework[2], we only achieve a sample rate of 5 Hz for our webcam-based experiments. As changes of fixations and saccades usually happen within the range of 200 ms to 400 ms [18], reliable gaze

[2] It is based on an iterative algorithm that each detection runs after the previous detection is finished.

Table 1: Features leveraged in the detection of participants' mind-wandering

Feature Name	**Explanation**
	Global Features
Fixation Duration	the durations (ms) of fixations
Saccade Duration	the durations (ms) of saccades
Saccade Distance	the distances (pixel) of saccades
Saccade Angle	the angles (degree) between saccades and the horizon
Number of Saccade	total number of saccades
Horizontal Saccade Ratio	the proportion of the number of saccades which have saccade angles less than 30 degree
Fixation Saccade Ration	the ratio of the durations of fixations to the duration of saccades
	Local Features
Saccade Landing	the proportion of the number of saccades landing in different areas
Fixation Duration AOI	the durations (ms) of fixations located in different areas

data comparable to the one provided by the high-speed Tobii tracker is impossible to obtain using such a low sample rate and thus needs to be estimated algorithmically. For this purpose we implement micro-saccade detection as discussed in [6]: we first determine whether the movement between two consecutive gaze points is a saccade based on the movements' velocity. Then we treat gaze points between two saccades as a fixation. If there is only a single gaze point between two saccades, we assume this gaze point is a fixation with a duration between this gaze point and the previous gaze point. After the detection of saccades and fixations, we can generate the same 58 features as already shown in Tab. 1. Intuitively, the feature vectors from the webcam-based solution are less precise (as the sampling rate is much lower), however, we will show later that they still show comparable classification performance as we aggregate features over the time spans between consecutive bells, thus this imprecision carries little weight.

To train our classifiers, we adopt leave-one-participant-out cross-validation [13]. In each run, the data of one participant is selected as test data and the data of all other participants is used for training. Based on the results reported in previous works [1, 2, 13], the collected data on learners' mind-wandering is usually unbalanced with considerably less than 50% of probes resulting in reported mind-wandering. We counter the effects of this imbalance by applying the oversampling method Synthetic Minority Over-sampling Technique (SMOTE) [5].

We have two requirements for our choice of classifiers as follows:

1. The selected models trained with our data can be used effectively to infer mind-wandering in data of unseen participants.
2. The selected models trained with our data can be used in *real-time* mind-wandering detection.

For the first requirement, we consider the bias-variance trade-off of machine learning models and the data size in our experiments. We select Logistic Regression, Linear SVM and Naive Bayes classifiers in our experiments as they have a low variance on small datasets like ours. These classifiers are also suitable for our second requirement. Since the trained models are small and require few inference steps, they can easily be integrated into Web applications within MOOC platforms.

In order to determine the effect of different feature types, we evaluate different subsets of features in our experiments: (i) global features only (G), (ii) local features only (L) and (iii) the combination of global and local features (G+L). Since we also include SMOTE as a pre-processing step to deal with the unbalanced nature of our data, overall we report results on six different setups.

4 Results

In this section, we focus on the experimental results of our study and described mind-wandering detection methods. We address two main research questions:

RQ1: How many mind-wandering reports are collected from participants across each video, and what can be learned from them?

RQ2: How well does our webcam-based mind-wandering detection method perform, and how does it compare to detection based on data collected from a professional eye tracker?

For **RQ2**, we first compare the overall effectiveness of our three selected classifiers with different sets of gaze features. Then, we delve deeper into the mind-wandering detection results. Considering that the mind-wandering reports are not evenly distributed among participants nor across the entire length of the lecture videos, we address two sub-questions **RQ2.1** and **RQ2.2**. A final sub-question is dedicated to the generalizability of our trained models.

RQ2.1: Does mind-wandering detection perform equally well across all participants?

RQ2.2: Does mind-wandering detection perform equally well across the entire length of a lecture video?

RQ2.3: Does a mind-wandering detection model trained on one video perform well to detect mind-wandering on a different video?

4.1 Exploratory Analysis of Mind-Wandering Reports

In order to answer **RQ1**, we now analyze our participants' mind-wandering behaviour while watching the two MOOC lecture videos.

In Fig. 4, the distributions of participants' reported mind-wandering events over the course of each of the two videos are shown. As discussed in the last section, participants were shown both videos in a random order, which is also reflected in the diagram. As the number of participants in each of the experimental groups is very small, no statistically significant conclusions can be drawn. However, it is visible that mind-wandering is indeed a rather frequent occurrence even for very short video lectures of roughly 7 minutes: our measured mind-wandering rate is 29%; i.e. in 71% of all bell time spans, our subjects actually stayed focused. In addition, it appears that our participants tire considerably during the second video when the experiment draws to its conclusion. This feedback was pro-actively provided by several of our participants in a post-experiment questionnaire, and seems to be at least anecdotally confirmed by the presented mind-wandering reports.

Table 2: Mind-wandering detection results based on gaze data (G means global features and L means local features)

Data	*Feature*	*SMOTE*	*Precision*	*Recall*	*F1*	*Classifier*
Baseline	–	–	0.290	0.291	0.290	–
Tobii Data	G	–	0.316	0.515	0.350	Logistic Regression
	G	✓	**0.358**	0.487	0.336	Logistic Regression
	L	–	0.263	0.625	0.309	Logistic Regression
	L	✓	0.294	**0.682**	**0.364**	Naive Bayes
	G+L	–	0.342	0.486	0.335	Naive Bayes
	G+L	✓	0.346	0.502	0.330	Linear SVM
WebGazer Data	G	–	0.309	0.671	0.395	Naive Bayes
	G	✓	0.306	**0.744**	**0.405**	Naive Bayes
	L	–	0.313	0.650	0.394	Naive Bayes
	L	✓	**0.320**	0.691	0.403	Naive Bayes
	G+L	–	0.289	0.696	0.378	Naive Bayes
	G+L	✓	0.286	0.674	0.378	Naive Bayes

4.2 Mind-Wandering Detection

In order to answer **RQ2**, we investigate how accurately we can detect participants' mind-wandering based on gaze data extracted by `WebGazer.js` compared to Tobii's `X2-30`. The results are shown in Tab. 2. The results are based on the nested leave-one-participant-out cross-validation, which means that a leave-one-participant-out cross-validation is used as the inner cross-validation for model selection and a leave-one-participant-out cross-validation is used as the outer cross-validation for measuring performance of the selected model. For the sake of brevity[3], we only list the best performing classifier for each feature set. As a *baseline method*, we used a random classifier which includes the knowledge that the mind-wandering rate is 0.29 and thus each feature vector is labeled as mind-wandering with a probability of 0.29. Since accuracy is not a suitable metric for unbalanced data, precision, recall, and F1-measure are reported.

Based on our results in Tab. 2, all our methods are significantly better than the random baseline according to all three metrics. We do not observe a large

[3] The full results, as well as all hyperparameter settings of the classifiers can be found online [25].

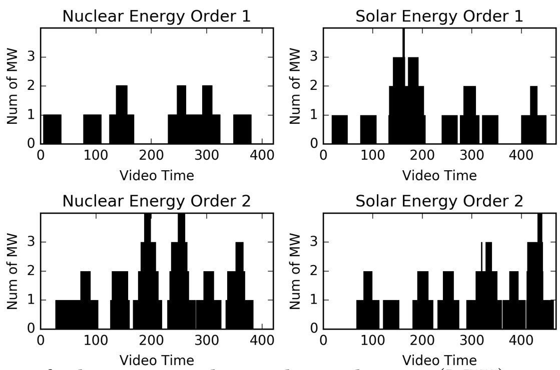


Fig. 4: Overview of the reported mind-wandering (MW) reports across the MOOC videos. Due to the randomized video order in the experiment, we partitioned the results according to whether the video was shown first ("order 1") or last ("order 2"). The video time displays the number of seconds since the start of the video.

Table 3: Statistics of detection results on individual participants ($P_{highest}$ shows the detection results of the participant with highest F1-measure, P_{lowest} with lowest)

Data	*Metrics*	*Max*	*Min*	*Mean*	*Std*	$P_{highest}$	P_{lowest}
Tobii	Precision	0.714	0	0.294	0.198	0.600	0
Data	Recall	1.000	0	0.682	0.357	0.857	0
	F1	0.706	0	0.364	0.200	0.706	0
WebGazer	Precision	0.700	0	0.306	0.209	0.700	0
Data	Recall	1.000	0	0.744	0.354	1.000	0
	F1	0.824	0	0.405	0.244	0.824	0

impact of SMOTE: applying the SMOTE pre-processing method on Tobii data slightly increases *Precision*, however it has no effect on the detection results on `Webgazer.js` data. The combination of local and global features does not benefit the detection on Tobii data nor the detection on `Webgazer.js` data.

All our reported F1 scores are slightly lower than reported by previous research [2] which relied on similar features and classifiers. We believe the difference (0.1 in F1 score) to be due to the slightly different data collection setup: Bixler et al. [2] utilized a short movie instead of MOOC lectures and free self-reporting instead of periodic self-reporting to obtain mind-wandering reports. With respect to the evaluated classification methods, we find that the Gaussian Naive Bayes models outperform the other approaches on `WebGazer.js` data in every feature set combination.

The most surprising finding in this experiment is that compared to the Tobii data we achieve higher *Recall* and *F1* scores based on the gaze features extracted from `WebGazer.js` data. Based on our intuition, features extracted from the data which is generated from the high-quality eye tracker `X2-30` should lead to a more accurate detection of mind-wandering, than features extracted from the data which is generated by a standard webcam. A possible reason for this experimental artifact is the small number of participants in our study; in future work we plan increase our participant pool to at least 100 participants.

Based on Tab. 2, we now delve deeper into our mind-wandering detection results. In order to answer **RQ2.1**, we investigate the detection results on each participant separately. For this step, we select the best-performing models for each data source (Tab. 2). For the detection on Tobii data, we use Gaussian Naive Bayes with local features and the SMOTE method. For the detection on `Webgazer.js` data, we use Gaussian Naive Bayes with global features and the SMOTE method. The results are shown in Tab. 3. We observe that across all metrics, the minimum observed accuracy is zero (for both Tobii and Webgazer data), which implies that there are participants for whom our prediction is not working at all. At the same time, we observe that at best a participant's mind-wandering can be detected with high accuracy with an F1 of 0.7 (Tobii data) and 0.8 (Webgazer data) respectively. The large standard deviations across the three metrics - 0.2 to 0.35 - further show that the accuracy of our detector varies widely between participants. Therefore, we conclude that *the detection does not work equally well for all participants in our experiments.*

Table 4: Detection across the entire length of the video (*Part 1* means the first half part of the video, and *Part 2* means the second half part of the video)

Data	*Metrics*	*Solar Energy* *Part 1*	*Part 2*	*Nuclear Energy* *Part 1*	*Part 2*
Tobii Data	Precision	0.147	0.410	0.276	0.321
	Recall	0.308	0.763	0.397	0.462
	F1	0.195	0.474	0.285	0.369
WebGazer Data	Precision	0.365	0.240	0.295	0.327
	Recall	0.615	0.500	0.462	0.615
	F1	0.438	0.285	0.344	0.416

Table 5: Detection with model translation (i.e. using a model on a different video than it was trained on)

Data	*Metrics*	*Trained on Solar* *used in Solar*	*used in Nuclear*	*Trained on Nuclear* *used in Nuclear*	*used in Solar*
Tobii Data	Precision	0.267	0.171	0.294	0.149
	Recall	0.705	0.372	0.410	0.205
	F1	0.355	0.229	0.296	0.150
WebGazer Data	Precision	0.240	0.298	0.346	0.344
	Recall	0.679	0.692	0.596	0.667
	F1	0.317	0.401	0.392	0.423

Based on the analysis in Sec. 4.1, we find that mind-wandering is not evenly distributed throughout a video. This leads to our **RQ2.2**. We split each video into two parts with the same length. Then, for each part of the video, we use the data of the other part and the data of the other video to train the model and to detect the mind-wandering in this specific left-out part of the video. The models, feature sets and the SMOTE method used in this experiments are same as in **RQ2.1**. The results are shown in Tab. 4. We conclude that *the detection of mind-wandering cannot be made equally well across the entire length of the lecture videos in our experiments*. For `X2-30` data, we find the results of the mind-wandering detection in the second part of the same video to be much better than the first part. For `WebGazer.js` data, we observe no trend, the results vary depending on the lecture video. We hypothesize this result to be connected to the fact that different participants were shown the videos in different orders.

Our last experiment answers **RQ2.3**. So far we have shown that our method can detect a participant's mind-wandering based on a model trained on the gaze data and mind-wandering reports of other participants. To scale out, we need to determine to what extent we can detect learners' mind-wandering in video lectures of one course with a model trained in lecture videos of other courses. If we were to obtain good detection results for such scenarios, there may be a general model which can be used in different lecture videos at scale (i.e., "train once, deploy everywhere"). In this experiment, the experimental settings for classifiers, feature sets and the SMOTE method on different kinds of data are same as in our previous experiments (**RQ2.1** and **RQ2.2**). We evaluate the cross-video performance by training our model on one video, and test the performance of the model using the other video. The results of all video combinations are shown in Tab. 5. For reference, this table also includes training and testing using the same video, using leave-one-participant-out cross-validation.

Based on the results in Tab. 5, we find the model trained on `WebGazer.js` data to be more robust to a change of video context than the model trained on `X2-30` data. We also observe that it does matter whether we train on video A and test on B or vice versa as results are comparable. Overall, we believe that *a model trained on the `WebGazer.js` data collected on one video can lead to good predictions in other videos*, at least if the videos share similarities with respect to style and type as in our scenario.

5 Conclusions

In this paper, we presented a study on the automatic and scalable detection of mind-wandering during lecture video watching collected by a standard consumer-grade webcam. In a lab study we compared the effectiveness of a webcam plus the open-source library `WebGazer.js` to the effectiveness of the specialized (and expensive) `Tobii X2-30` for the task of mind-wandering detection. In our experiments, we could show that the accuracy of our webcam-based approach is on par with the specialized eye-tracking device. This opens the way for large-scale experiments in real-world MOOCs, allowing for both investigating learners' mind-wandering behavior and investigating the effectiveness of interventions based on mind-wandering detection in future research under realistic conditions.

Our work is in a preliminary stage and has a number of limitations including the small pool of participants all sharing similar educational backgrounds. Similarly, the number of evaluated MOOC videos is very limited and both videos have a comparable (but very common) style. Thus, it is unclear how well our approach can be applied to completely different types of videos or user groups. In addition, we relied on a number of established and straightforward-to-implement features; we expect a further boost in detection accuracy when more sophisticated features are introduced.

A core contribution provided by our work is the published repository of data collected during our controlled lab study. In addition to including the mind-wandering reports of our experiment's participants, we also provide the full set of gaze data obtained by the `X2-30` and our webcam as well as the complete results of our data analysis.

References

1. R. Bixler and S. DMello. Toward fully automated person-independent detection of mind wandering. In *UMAP'14*, pages 37–48. Springer, 2014.
2. R. Bixler and S. DMello. Automatic gaze-based user-independent detection of mind wandering during computerized reading. *User Modeling and User-Adapted Interaction*, 26(1):33–68, 2016.
3. D. M. Bunce, E. A. Flens, and K. Y. Neiles. How long can students pay attention in class? a study of student attention decline using clickers. *Journal of Chemical Education*, 87(12):1438–1443, 2010.
4. P. Cameron and D. Giuntoli. Consciousness sampling in the college classroom or is anybody listening?. *Intellect*, 101(2343):63–4, 1972.

5. N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer. Smote: synthetic minority over-sampling technique. *Journal of artificial intelligence research*, 16:321–357, 2002.
6. R. Engbert and R. Kliegl. Microsaccades uncover the orientation of covert attention. *Vision research*, 43(9):1035–1045, 2003.
7. A. H. Johnstone and F. Percival. Attention breaks in lectures. *Education in chemistry*, 13(2):49–50, 1976.
8. M. A. Killingsworth and D. T. Gilbert. A wandering mind is an unhappy mind. *Science*, 330(6006):932–932, 2010.
9. S. I. Lindquist and J. P. McLean. Daydreaming and its correlates in an educational environment. *Learning and Individual Differences*, 21(2):158–167, 2011.
10. K. K. Loh, B. Z. H. Tan, and S. W. H. Lim. Media multitasking predicts video-recorded lecture learning performance through mind wandering tendencies. *Computers in Human Behavior*, 63:943–947, 2016.
11. J. McLeish. *The lecture method*. Cambridge Institute of Education, 1968.
12. R. McMillan, S. B. Kaufman, and J. L. Singer. Ode to positive constructive daydreaming. *Frontiers in psychology*, 4:626, 2013.
13. C. Mills, R. Bixler, X. Wang, and S. K. DMello. Automatic gaze-based detection of mind wandering during narrative film comprehension. In *EDM'16*, pages 30–37, 2016.
14. A. Papoutsaki, N. Daskalova, P. Sangkloy, J. Huang, J. Laskey, and J. Hays. Webgazer: scalable webcam eye tracking using user interactions. In *IJCAI'16*, pages 3839–3845, 2016.
15. K. Rayner. Eye movements in reading and information processing: 20 years of research. *Psychological bulletin*, 124(3):372–422, 1998.
16. E. F. Risko, N. Anderson, A. Sarwal, M. Engelhardt, and A. Kingstone. Everyday attention: variation in mind wandering and memory in a lecture. *Applied Cognitive Psychology*, 26(2):234–242, 2012.
17. O. Rodriguez. The concept of openness behind c and x-MOOCs (Massive Open Online Courses). *Open Praxis*, 5(1):67–73, 2013.
18. D. D. Salvucci and J. H. Goldberg. Identifying fixations and saccades in eye-tracking protocols. In *Proceedings of the 2000 Symposium on Eye Tracking Research & Applications*, pages 71–78. ACM, 2000.
19. J. Smallwood, D. J. Fishman, and J. W. Schooler. Counting the cost of an absent mind: Mind wandering as an underrecognized influence on educational performance. *Psychonomic bulletin & review*, 14(2):230–236, 2007.
20. J. Smallwood and J. W. Schooler. The restless mind. *Psychological bulletin*, 132(6):946–958, 2006.
21. J. Stuart and R. Rutherford. Medical student concentration during lectures. *The lancet*, 312(8088):514–516, 1978.
22. K. K. Szpunar, N. Y. Khan, and D. L. Schacter. Interpolated memory tests reduce mind wandering and improve learning of online lectures. *Proceedings of the National Academy of Sciences*, 110(16):6313–6317, 2013.
23. T. Tan, H. Zou, C. Chen, and J. Luo. Mind wandering and the incubation effect in insight problem solving. *Creativity Research Journal*, 27(4):375–382, 2015.
24. K. Wilson and J. H. Korn. Attention during lectures: Beyond ten minutes. *Teaching of Psychology*, 34(2):85–89, 2007.
25. Y. Zhao, C. Lofi, and C. Hauff. EC-TEL 2017 companion webpage. `https://yue-zhao.github.io/MWDET_Project/`, 2017.