IntelliEye: Enhancing MOOC Learners' Video Watching Experience through Real-Time Attention Tracking

Tarmo Robal
Tallinn University of Technology
Tallinn, Estonia
tarmo.robal@ttu.ee

Yue Zhao, Christoph Lofi, Claudia Hauff*
Delft University of Technology
Delft, Netherlands
{y.zhao-1,c.lofi,c.hauff}@tudelft.nl

ABSTRACT

Massive Open Online Courses (MOOCs) have become an attractive opportunity for people around the world to gain knowledge and skills. Despite the initial enthusiasm of the first wave of MOOCs and the subsequent research efforts, MOOCs today suffer from retention issues: many MOOC learners start but do not finish. A main culprit is the lack of oversight and directions: learners need to be skilled in self-regulated learning to monitor themselves and their progress, keep their focus and plan their learning. Many learners lack such skills and as a consequence do not succeed in their chosen MOOC. Many of today's MOOCs are centered around video lectures, which provide ample opportunities for learners to become distracted and lose their attention without realizing it. If we were able to detect learners' loss of attention in real-time, we would be able to intervene and ideally return learners' attention to the video. This is the scenario we investigate: we designed a privacy-aware system (`IntelliEye`) that makes use of learners' Webcam feeds to determine—in real-time—when they no longer pay attention to the lecture videos. `IntelliEye` makes learners aware of their attention loss via visual and auditory cues. We deployed `IntelliEye` in a MOOC across a period of 74 days and explore to what extent MOOC learners accept it as part of their learning and to what extent it influences learners' behaviour. `IntelliEye` is open-sourced at https://github.com/Yue-ZHAO/IntelliEye.

KEYWORDS

MOOCs; online learning; IntelliEye

ACM Reference Format:
Tarmo Robal and Yue Zhao, Christoph Lofi, Claudia Hauff. 2018. IntelliEye: Enhancing MOOC Learners' Video Watching Experience through Real-Time Attention Tracking. In *HT'18: 29th ACM Conference on Hypertext & Social Media, July 9–12, 2018, Baltimore, MD, USA.* ACM, New York, NY, USA, 9 pages. https://doi.org/10.1145/3209542.3209547

*The first and second author of this paper contributed equally.

HT'18, July 9–12, 2018, Baltimore, MD, USA

ACM ISBN 978-1-4503-5427-1/18/07…$15.00
https://doi.org/10.1145/3209542.3209547

1 INTRODUCTION

In 2011, the *MOOC revolution* began: Stanford University offered the first MOOC on Artificial Intelligence followed by more than 160,000 learners worldwide. The idea of MOOCs quickly spread. A major motivation behind MOOCs is the provision of ubiquitous learning to people from all walks of live. Today, MOOCs are being offered by many world-renowned universities on platforms such as Coursera, FutureLearn and edX[1], reaching millions of learners. At the same time, the initial predictions of this revolution have not come to pass—MOOCs today suffer from a lack of retention with usually less than 10% (in extreme cases < 1%) of learners succeeding [9]. Examining the current nature of MOOCs reveals an important clue as to why they fail to realize their potential: although they offer flexibility, and scale, they do not involve truly novel technologies. Most MOOCs today revolve around a large number of videos[2] and automatically graded quizzes and little else. This setup (largely chosen for its inherent scalability), requires learners to be skilled in *self-regulated learning* [21], that is, to monitor themselves and their progress, keep their focus and plan their learning. Many learners lack such skills and as a consequence do not succeed. In this paper we present `IntelliEye`, a system we designed to directly tackle the "loss of focus" issue during MOOC lecture video watching by detecting it in real-time and alerting the learner to it. We focus our efforts on the video watching activity as (i) learners spend a large portion of their time in a MOOC on it; (ii) learners are prone to lose their focus even in short lecture videos of six to ten minutes [30], a common video length in MOOCs; (iii) video watching is a rather passive activity which provides ample opportunities for learners to become distracted—and engage in "heavy media multitasking" [13] by reading their emails, surfing the Web, etc.—and lose their focus often without realizing it; and (iv) inattention has been shown to be significantly and negatively correlated with learning efficiency [26].

How exactly can we detect learners' loss of focus *in real-time* and *at scale*? How can we alert the learner to her loss of focus? One answer to these questions lies in the ubiquitous availability of Webcams in today's laptops: `IntelliEye` employs the Webcam feed to observe learners' activities during their time on the MOOC platform and intervenes (e.g. by delivering an auditory signal) if it detects a loss of focus. All of these actions are performed by `IntelliEye` in a *privacy-aware* manner: none of the data or computations leaves a user's machine. Prior works [1, 2, 18, 30] exploited eye-tracking to determine a user's attention state, though these studies were either conducted with commercial high-quality hardware eye-tracking devices and/or well-settled experimental lab conditions [30]. In contrast, in our work we make use of commonly available Webcams

[1] https://www.edx.org/
[2] The MOOC we deploy `IntelliEye` in contains 104 lecture videos.

and deploy `IntelliEye` "in the wild", to 2,612 MOOC learners in an actual MOOC, instead of a controlled lab study.

We conduct our analyses of `IntelliEye`'s use along three dimensions: (1) the **technological capabilities** of MOOC learners' hardware, (2) the **acceptance** of `IntelliEye` by MOOC learners, and, (3) the **effect** of `IntelliEye` on MOOC learners' behaviour. Specifically, we investigate the following research questions:

- **RQ1** To what extent is MOOC learners' hardware capable to enable the usage of technologically advanced widgets such as `IntelliEye`?
- **RQ2** To what extent do MOOC learners accept technology that is designed to aid their learning but at the same time is likely to be perceived as privacy-invading (even though it is not)? Are certain types of MOOC learners (e.g. young learners, or highly educated ones) more likely to accept this technology than others?
- **RQ3** What impact does `IntelliEye` have on learners' behaviours and actions? To what extent does `IntelliEye` affect learners' video watching behaviour?

Our main findings can be summarized as follows:

- We find that most learners (78%) use hardware and software setups which are capable to support such widgets, making the wide-spread adoption of our approach realistic from a technological point of view.
- The majority of learners (67%) with capable setups is reluctant to allow the use of Webcam-based attention tracking techniques, citing as main reasons privacy concerns and the lack of perceived usefulness of such a tool.
- Among the learners using `IntelliEye` we observe (i) high levels of inattention (on average one inattention episode occurs every 36 seconds—a significantly higher rate than reported in previous lab studies) and (ii) an adaptation of learners' behaviour towards the technology (learners in conditions that disturb the learner when inattention occurs exhibit fewer inattention episodes than learners in a condition that provides less disturbance).

2 RELATED WORK

Attention Loss in the Learning Process

Identifying and tracking learners' loss of attention in the classroom has been explored in a myriad of ways since the 1960s, including the analysis of students' notes [7, 15], the observation of inattention behaviors (by observers, stationed at the back of the classroom) [8], the retention of course content [16], probes (requiring participants to record their attention at particular given points in time) [12, 27] and self-reports (requiring participants to report when they become aware of their loss of attention) [3]. A common belief was that learners' attention decreases considerably after 10-15 minutes into the lecture [27]. Later, Wilson and Korn [29] challenged this claim and argued that more research is needed, a call picked up by Bunce et al. [3] who found that learners start losing their attention early on in higher-education lectures and may cycle through several attention states within 9-12 minute course segments.

With the advent of online learning, the issue of attention loss, how to measure it and how it compares to classroom attention lapses received renewed attention. Different studies have shown that in online learning environments (often simulated in lab settings where participants watch lecture videos), attention lapses may be even more frequent than in the classroom setting. Risko et al. [23] used three one hour video-recorded lectures with various topics (psychology, economics, and classics) in their experiments, probing participants four times throughout each video. The attention-loss frequency was found to be 43%. In addition, a significant negative correlation between test performance and loss of attention was found. Szpunar et al. [28] studied the impact of interpolated tests on learners' loss of attention within online lectures, asking participants to watch a 21-minute video lecture (4 segments with 5.5 minutes per segment) and report their loss of attention in response to random probes (one per segment). In their study, the loss of attention frequency was about 40%. Loh et al. [13] also applied probes to measure learners' loss of attention, finding a positive correlation between media multitasking activity and learners' loss of attention (average frequency of 32%) whilst watching video lectures. Based on these considerably high loss of attention frequencies we conclude that reducing loss of attention in online learning is an important approach to improve learning outcomes.

Automatic Detection of Attention Loss

Inspired by the eye-mind link effect [22], a number of previous studies [1, 2, 18] focused on the automatic detection of learners' loss of attention by means of gaze data. In [1, 2], Bixler and D'Mello investigated the detection of learners' loss of attention during computerized reading. To generate the ground truth, the study participants were asked to manually report their loss of attention when an auditory probe (i.e. a beep) was triggered. Based on those reports, the loss of attention frequency ranged from 24% to 30%. During the experiment, gaze data was collected using a dedicated eye-tracker. Mills et al. [18] asked study participants to watch a 32 minute, non-educational movie and self-report their loss of attention throughout. In order to detect loss of attention 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 in. Zhao et al. [30] presented a method to detect inattention similar to the studies in [18], but optimized for a MOOC setting (including the use of a Webcam alongside a high-quality eye-tracker). All mentioned approaches relying on the eye-mind link share two common issues: (i) they are usually unable to provide real-time feedback as they are trained on eye-gaze recordings with sparse manually provided labels (e.g., most approaches have a label frequency of 30-60 seconds, which directly translates into a detection delay of similar length), and (ii) the reported accuracy is too low for practical application (e.g., [30] reports detection accuracy between 14%-35%). Lastly, we note that besides the eye-mind link, another recent direction is the use of heart rate data (measured for instance by tracking fingertip transparency changes [20]) to infer learners' attention.

MOOC Interventions

We now discuss MOOC interventions, especially those geared towards video watching and towards improving self-regulated learning. Existing research on MOOC videos is largely concerned with the question of what makes a MOOC video engaging and attractive to learners; examples include the overlay of an instructor's face over the lecture slides [10], shorter video segments instead of one long lecture video [6], and the overlay of an instructor's gaze to enable learners to more easily follow the video content [25].

Few works have considered the issue of self-regulated learning in MOOCs, largely because this requires approaches that are personalized and reactive towards each individual learner. Simply informing learners about the best strategies for self-regulated learning at the beginning of a MOOC is not sufficient [11]. Davis et al. [5] recently designed a visual "personalized feedback system" that enables learners to learn how well they are doing compared to successful passers from a previous MOOC edition (in terms of time spent on the platform, their summative assessment scores and so on). This comparison, even though this feedback moment was rare (once a week), enabled learners to self-regulate their learning better, leading to significantly higher completion rates for learners exposed to the feedback system. A prior study by Davis et al. [4] had indicated that non-compliance among learners is a difficult obstacle in very simple interventions: the authors had included an extra question in each week of a MOOC, asking learners to write about their study plans (and thus make learners think about those plans). Few learners saw the benefit of this question (it was ungraded) and thus very few complied.

Overall, we have shown that attention lapses are a regular occurrence in the classroom and occur with even greater frequency in online learning, where learners are prone to digital multitasking. We have also presented some drawbacks of sophisticated eyetracking-based attention loss detectors (accuracy and timeliness of detection) and finally we have pointed out the difficulty of bringing self-regulated learning into the MOOC scenario due to learners' non-compliance. In response to these findings we have designed `IntelliEye`, a robust attention loss (by using face detection as a proxy) detector that requires no additional actions by the learners beyond what they usually do on a MOOC platform, provides personalized feedback, is privacy-aware and detects a loss of attention in near-real-time (with at most 2 seconds delay).

3 INTELLIEYE

3.1 Architecture

The goal of `IntelliEye` is to provide real-time feedback on learner's attention, and is based on a set of heuristics *reliably implementable* on a wide variety of hardware setups: (1) if the browser tab/window containing the lecture video is not visible to the learner, `IntelliEye` triggers an inattention event; (2) we assume a learner is inattentive if her face cannot be detected for a period of time, i.e. we employ face tracking as a robust proxy of attention tracking[3]; (3) if the face tracking module detects a loss of the face we consider the mouse movements as a safety check: if no face is detected but the mouse is being moved, no event is triggered.

The resulting high-level architecture is shown in Figure 1. We implemented `IntelliEye` in JavaScript, as the edX platform allows custom JavaScript to be embedded in course modules—thus providing us with an easy way to "ship" `IntelliEye` to all learners in our MOOC. As visible in Figure 1, `IntelliEye` resides exclusively on the client to ensure learners' privacy; usage logs are send to our dedicated `IntelliEye` log server for the purpose of evaluating `IntelliEye`, though this communication is not necessary for `IntelliEye` to function. This setup requires `IntelliEye` to be light-weight and resource-saving as all computations are carried out on the learner's device and within the resource limits of a common Web browser. We now describe the seven architecture modules that `IntelliEye` consists off.

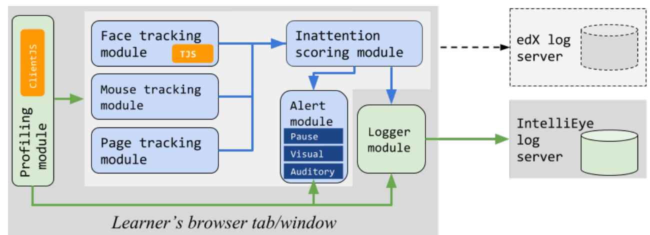

Figure 1: `IntelliEye`'s high-level architecture. The profiling and logger modules are always active; the attention tracking and alerting modules are only enabled if supported setup is detected and learner has granted access to Webcam feed.

3.1.1 *Profiling Module.*

In order to provide a smooth user experience for MOOC learners we limit the full usage of `IntelliEye` to devices that fulfill certain device setup requirements, a situation we call *supported setup*. We rely on the `ClientJS`[4] library to determine the device type, operating system and browser version of the learner's device and activate the inattention tracking modules only if a supported setup is detected. The requirements are as follows:

(1) The device is not a mobile device and is not running iOS or Android, due to their incompatibility with `IntelliEye`.
(2) The browser used is either: Chrome 54+ (i.e. version 54 or higher), Firefox 45+ or Opera 41+ to ensure the availability of JavaScript dependencies necessary for `IntelliEye`.
(3) The device has at least one usable Webcam as detected via the Media Capture and Streams API.

If the profiling yields an *unsupported setup*, a log entry is sent to our `IntelliEye` log server and no further modules are activated.

The profiling module is also responsible for extracting the learner's edX user ID, which in turn determines which alert type the learner receives in our experiments.

3.1.2 *Face Tracking Module.*

In `IntelliEye`, we use face tracking to proxy inattention detection, thus aiming at overcoming the reported shortcomings of gaze tracking with respect to response time and reliability: if a learner's face is not visible in front of the screen when a lecture video is playing, we argue that she is likely not paying attention.

We initially experimented with two open-source libraries for this purpose—`WebGazer.js`[5] and `tracking.js` [14][6] (or `TJS` for short)—and investigated their suitability for Webcam-based inattention detection using face tracking in a user study with 20 participants [24]. As an upper bound, we also included the high-end

[3]We note that this is a lower-bound for inattention, as learners watching the video may still not pay attention.

[4]https://github.com/jackspirou/clientjs
[5]https://webgazer.cs.brown.edu
[6]https://trackingjs.com

hardware eye-tracker `Tobii X2-30 Compact`. We evaluated all three setups using fifty behaviours that learners typically execute in front of their computer; thirty-five of those behaviours should lead to a face detection loss (such as *Check your phone*; *Look right for 10 seconds*) and fifteen should not (e.g. *Reposition yourself in the chair*; *Scratch the top of your head*). During the study the participants were asked to perform each of the fifty tasks in turn. The study showed that only `TJS` has a competitive accuracy: it is able to detect 77.8% (compared to `Webgazer.js`'s 14.8%) of the face hit/face miss behaviours that the `Tobii X2-30 Compact` was identifying correctly. We also measured the delay in detecting inattention, i.e. the difference in seconds between the behaviour being performed by a study participant and the inattention being detected: $0.6 \pm 1.1s$ for `TJS` and $1.3 \pm 1.0s$ for `Webgazer.js`. Based on these results, we chose `TJS` as our face detection library.

The module performs face presence detection (via `TJS`) from the Webcam feed every 250ms and reports a boolean (face present or absent) to the *Inattention scoring module*. We chose this time interval not to overburden the computational resources of the learner's device.

3.1.3 Mouse Tracking Module. This module acts as a sanity check for the face tracking module: if the face tracking module reports loss of a face and the learner is still moving the mouse in the active MOOC window, we assume that the face tracking module misclassified the situation and do not raise an inattention alert. This module tracks the absence or presence of mouse movements every 250ms and reports it to the *Inattention scoring module*.

3.1.4 Page Tracking Module. This module tracks the visibility of the browser window or tab that contains the edX page (and thus the lecture video) using the `document.hidden()` Web API call. A value is produced every 250ms and forwarded to the *Inattention scoring module*.

3.1.5 Inattention Scoring Module. This module estimates inattention of a learner by aggregating the data obtained from the tracking modules based on the heuristics already introduced at the start of § 3.1: a learner is inattentive if her face is not trackable unless there is mouse movement and the video player browser window is visible. The input from the three scoring modules is aggregated over a sliding time window of five seconds—we chose this time window based on our user study with 50 typical activities during MOOC video watching, where we found the longest activity to take approximately five seconds. Recall that each module has a fixed sampling rate of 250 ms, and thus our sliding window takes into account 20 measurement points from each tracking module.

More formally, the input to this module are the boolean values (i) for face presence $\mathcal{F} = (..., f_{n-20}, f_{n-19}, ..., f_n)$, (ii) mouse movement $\mathcal{M} = (..., m_{n-20}, m_{n-19}, ..., m_n)$, and (iii) page visibility $\mathcal{V} = (..., v_{n-20}, v_{n-19}, ..., v_n)$. To conserve computational resources, the module computes the attention state once a second. Algorithm 1 outlines the inattention decision process employed by the *Inattention Scoring module*. In essence, a weighted score is computed for the face presence and mouse movement values (lines 3 & 4), giving higher weights to more recent values. The visibility score of the video window is simply the last recorded value (line 5). Lines 6-9 compute face-tracking trends over time. The role of the face-tracking trend computation is to minimize the volume of false positives driven by learner behaviour, in particular sudden movements, bad position in front of the Webcam, or a temporary short time failure of `TJS` in detecting the face in Webcam video feed. Lines 10-11 show the rules the module employs to determine inattention based on the predefined threshold (which represents the minimum accepted score that is considered as attention, in our case $\mathcal{L} = 2.92$), computed scores and the trend. The threshold and rules are another outcome of our user study—they led to the highest accuracy in distinguishing between attention and inattention behaviours [24].

Algorithm 1 Inattention detection mechanism in `IntelliEye`

Require: $\mathcal{F}, \mathcal{M}, \mathcal{V}, \mathcal{L}$ — threshold value, $\mathcal{S}$—scores for $\mathcal{F}, \mathcal{M}, \mathcal{V}$
$\mathcal{T} = (t_1, t_2, ..., t_k)$ score queue of the trending functionality;
1: $inAttention \leftarrow False$
2: $n \leftarrow 20$
3: $S_F \leftarrow \sum_i f_{n-i}(n-i)/n$
4: $S_M \leftarrow \sum_i m_{n-i}(n-i)/n$
5: $S_V \leftarrow v_n$
6: $trend_F \leftarrow 0$
7: $T.dequeue(t_1); T.enqueue(t_k \leftarrow S_F)$
8: $(t_k > t_{k-1}) \Rightarrow trend_F \leftarrow 1$
9: $(t_k < t_{k-1}) \wedge (t_{k-1} < t_{k-2}) \Rightarrow trend_F \leftarrow -1$
10: $Q \leftarrow (S_F < L \wedge trend_F < 1)$
11: $(Q \wedge S_M < L \wedge S_V) \vee (Q \wedge \neg S_V) \vee (S_F > L \wedge \neg S_V) \Rightarrow inAttention \leftarrow True$

Note that the level of thresholding ($\mathcal{L}$) determines the sensitivity of `IntelliEye`—lowering the value will make the system less rigorous, increasing this value will on the other hand increase system responsiveness to learner behaviour.

3.1.6 Alert Module. We explored three different mechanisms—with varying levels of disruption—to raise learners' awareness about their detected loss of attention; none of these requiring an action from the user beyond returning their attention to the video at hand. In our experiment each learner is assigned to a single alert type, depending on their edX user ID detected by the *Profiling module*.

Pausing the video: When attention loss is detected `IntelliEye` will pause the currently playing lecture video. Once `IntelliEye` detects re-gained attention on the video, playing is resumed. At what position playing is resumed depends on *how long* the learner was not paying attention since pausing. The video is rewound to between 0 and 10 seconds before the attention loss was detected; we define three different configurations: (i) if the inattention period is less than 1.5 seconds, the video continues from where it was paused as it would be annoying for a learner to review content just seen and available in her short-time memory, but also to avoid repetitive 'rewind-and-play' situations; (ii) if the inattention lasted more than 10 seconds, the video is rewound 10 seconds which is the approximate lower level of human short-time memory (reported in between 10-30 seconds [17, 19]); and (iii) in all other cases it is rewound 3 seconds—rewind a little for rapid recall in case of distraction. This scheme ensures that the video will restart at a familiar point for the learner. The drawback of this mechanism is the severity of false alerts as the video will pause and thus the learner is disturbed if inattention was falsely determined.

Auditory alert: In this setup, the video keeps playing but an additional sound effect (a bell ring) is played repeatedly as long as inattention is detected. This setup is not as "annoying" as falsely pausing the video, but can still substantially disturb the learner.

Visual alert: In this version, `IntelliEye` visually alerts the learner by repeatedly flashing a red border around the video as long as inattention is detected. Figure 3 shows an example of this alert. This scenario is the least intrusive in case `IntelliEye` falsely detects inattention. It may also be the least effective, as learners who look away from the screen or minimize the browser tab/window will not be able to view the alert.

3.1.7 Logger Module. This module is responsible for logging `IntelliEye`'s usage. These logs are sent to our dedicated log server. Specifically, the following actions lead to logging (for log entries with categorical values we list all possible values within {...}):

Loading: When `IntelliEye` is loaded due to a learner accessing a course subsection[7] containing one or more video units we log `(timestamp, alertType {pause, visual, auditory}, userID, deviceSetup)`.

Video status change: Every change in the video's status (e.g. from paused to play) for a learner with supported setup leads to a log of the form `(videoID, timestamp, videoStatus {play, pause, seek, end}, videoTime, videoLength, videoSpeed, subtitles {on, off}, fullScreen {on, off})`. The `videoTime` entry refers to the point in time within the video the status changed.

`IntelliEye` status change: When a learner with a supported setup changes the status of `IntelliEye` (e.g. from disabled to enabled), we log `(videoID, timestamp, videoTime, videoLength, IntelliEyeStatus {allow, disallow, start, pause, resume, end})`. Information on the video is logged as most interactions with `IntelliEye` occur within the edX video player (cf. §3.2).

Inattention status change: This log event occurs when for a learner with a supported setup the attention status changes: `(videoID, timestamp, videoTime, videoLength, inattention {start, stop})`. Here, `start` indicates that inattention has been detected. The next event is generated when the status changes back to attention again (`stop`). As long as the inattention state is maintained, no further log events are generated.

Finally, we note that beyond the `IntelliEye` logs (cf. Figure 1), we also have access to the official edX logs, which contain information on all common actions learners perform within a MOOC on the edX platform such as quiz submissions, forum entries, clicks, views, and so on—data we use in some of our analyses.

3.2 User Interface

Having described `IntelliEye`'s architecture, we now turn to its user interface. Figure 2 shows `IntelliEye`'s welcome screen (potentially shown every time a MOOC learner opens a course subsection with one or more video units), describing its capabilities, and the positive impact it can have on learning. The learner has four choices: (i) to enable `IntelliEye` for this particular video only, (ii) to disable `IntelliEye` for this video only, (iii) to enable `IntelliEye` for all videos, and, (iv) to disable `IntelliEye` for all videos. If a learner opts for (iv), we ask the her for feedback on the decision (*"You have disabled IntelliEye. Please tell us why."*).

Once a learner enables `IntelliEye`, the face tracking module attempts to access the Webcam feed, which in all supported browsers triggers a dialogue controlled by the browser (*Will you allow edx.org to use your camera?*); once the learner chooses *Allow*, `IntelliEye` is fully functioning.

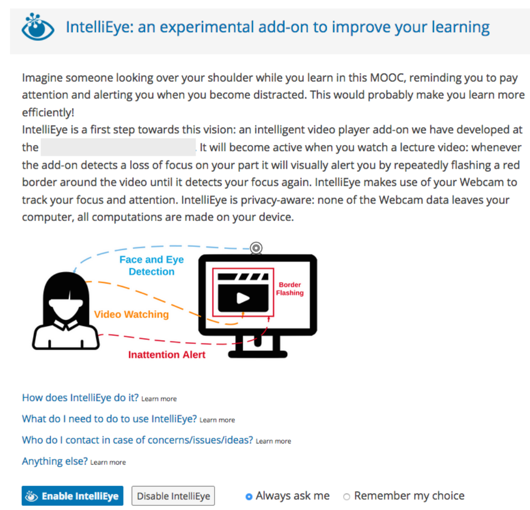

Figure 2: `IntelliEye` welcome screen.

Figure 3 shows how `IntelliEye` embeds itself in the edX video player. Here the learner can return to the welcome screen and change her enable/disable decisions (via the "eye" icon) and switch `IntelliEye` on or off on the fly. `IntelliEye`'s status is visible at all times: either 'Active' (`IntelliEye` is enabled, the video is not playing at the moment), 'Playing' (`IntelliEye` is enabled), or 'Not Active' (`IntelliEye` is disabled). Note that this change in the video player interface is only visible to learners with a supported setup. Learners on non-supported setups will receive the original edX video player without alterations.

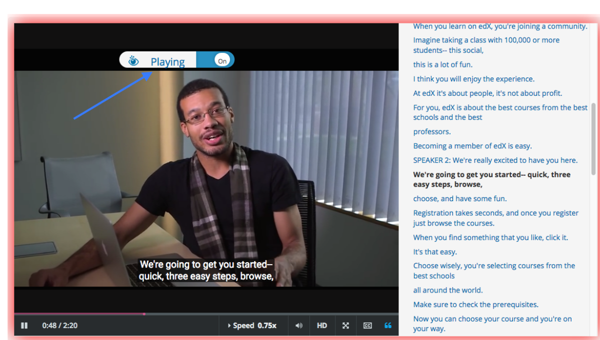

Figure 3: `IntelliEye`'s video player interface (arrow) embedded in the edX video player widget. The red hue around the video player is the visual alert we experiment with.

[7] A set of course elements semantically belonging together, cf. §4.

4 MOOC SETTING

We deployed `IntelliEye` in the MOOC *Introduction to Aeronautical Engineering (AE1110x)* offered by TU Delft on the edX platform. The MOOC's target population are learners who are looking for a first introduction to this particular field of engineering. The MOOC requires around 80-90 hours of work and consists of 104 videos and 332 automatically graded summative assessment questions. The MOOC is *self-paced*, that is, the MOOC is available for learners to enroll for up to 11 months. In contrast to the more common six to ten week MOOCs, learners can set their own schedule and their own pace. The MOOC was opened for enrollment on May 1, 2017 and remained so until March 31, 2018. `IntelliEye` was deployed for ten weeks (October 5, 2017 to December 17, 2017); it was available for all videos within the MOOC. A total of 2, 612 different learners visited the MOOC during the deployment period and were exposed to `IntelliEye`. We deployed `IntelliEye` in three different variants according to the manner of alerting learners to their lack of attention: video pause, auditory alert and visual alert (§ 3.1.6). We conducted an inter-subject study: each learner was randomly assigned (based on their learner ID) to one of the three conditions. Once assigned, a learner remained in that condition throughout the experiment. Table 1 shows the distribution of the 2, 612 learners across the three conditions.

Before turning to the analyses section, we introduce the relevant concepts and definitions:

Course subsection: on the edX platform, a course subsection refers to a sequence of course units (such as video units, quiz units and text units) that are grouped together, most likely because they all relate to the same topic. As an example, one of the subsections in our MOOC consists of the following sequence: video →video →text →quiz →video →quiz →text.

Session: refers to a sequence of logs from a single learner (active on a single device), with no more than thirty minutes time difference between consecutive log entries. This means that after thirty minutes of inactivity in the MOOC, we assume a new "learning" session starts (if the learner becomes active again). We combine the logs we retrieved from our `IntelliEye` log server with those collected by edX.

Supported session: refers to a session with a supported setup.

Unsupported session: a session without a supported setup.

Video session: a session in which at least one video was being played by the learner, regardless of the length of video playing.

`IntelliEye` session: refers to a supported session which is also a video session, and in which `IntelliEye` was running (which means that the learner did accept the terms of use and played a video while `IntelliEye` was active).

Non-`IntelliEye` session: a supported session which is also a video session, and in which `IntelliEye` is not active while the video was playing (this either means that the learner did not accept the terms of use, or manually disabled `IntelliEye`).

5 EMPIRICAL EVALUATION

5.1 RQ1: Technological Capabilities

The first question we consider is to what extent our MOOC learners (who, according to their edX profiles, hail from 138 different countries) have a supported device setup: according to Table 1, 78% of learners (across all three alert types) log in at least once with a device supported in `IntelliEye`. Among those 563 learners (22%) who never have a supported session, 223 of them only access the course with a mobile device (that is 9% of the overall learner population). If we drill down on the 340 learners with unsupported sessions on non-mobile devices, the most common reason is an outdated browser we do not support (e.g. Chrome 52, IE 11, Safari 10 and Safari 11), followed by the lack of a Webcam (in 118 cases). We do not observe a particular skew towards certain countries or regions; learners from India (104 learners) and learners from the US (93 learners) have the largest number of unsupported setups, which are also the two countries where most learners hail from (484 learners from India and 334 from the US).

Table 1: Learners exposed to `IntelliEye`. Shown is the number of learners: (i) in each alert type condition, (ii) with at least one session with supported setup, (iii) who used `IntelliEye` at least once, and (iv) not accepting `IntelliEye`.

Alert Types	#Exposed Learners	#Learners with 1+ Supported Sessions	#Learners with 1+ IntelliEye Session	#Learners without IntelliEye Session
Video pause	861	681	214	467
Auditory alert	902	703	208	495
Visual alert	849	665	236	429
Total	2612	2049	658	1391
% of total	–	78%	25%	53%

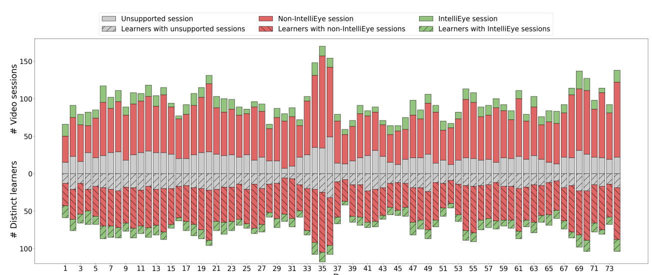

Figure 4: Distribution of video sessions and unique learners. A learner may be listed in more than one session type.

5.2 RQ2: Acceptance of `IntelliEye`

Having established that our hardware requirements are reasonable, we now turn to `IntelliEye`'s acceptance, i.e., are learners willing to enable a widget which observes them via a Webcam. As Table 1 shows, 32% of learners (658 out of 2049) with at least one supported session activate `IntelliEye` at least once.

We had two hypotheses on who engages with our intervention: (1) younger learners are more likely to engage than older ones, and (2) more active learners are more likely to engage than less active ones. To explore these hypotheses we computed various metrics for three different user groups (learners that do not engage with `IntelliEye`, learners that have one or two `IntelliEye` sessions and learners that have three or more `IntelliEye` sessions) as shown in Table 2[8]. We observe significant differences across

[8]Note that all our analyses consider the 74 days of `IntelliEye`'s deployment only, i.e. the number of sessions, the quiz scores, etc. are only computed for that time period.

Table 2: Learner attributes partitioned according to the use of `IntelliEye` (choices made on welcome page are not considered in grouping). Only learners with at least one supported video session are considered. $*$ indicates Student's t-test significance at $p < 0.05$ level. † and ‡ indicate Mann-Whitney U test significance at $p < 0.05$ and $p < 0.01$ levels respectively.

	Number of IntelliEye sessions		
	None	1-2	3+
Number of learners	1030	623	35
Median age	23	21^{*None}	22
Median prior education	Associate degree	High school ^{*None}	High school ^{*None}
Median av. session length (min)	27.77	27.44	$35.17^{\dagger None,1-2}$
Median #sessions	3	3	$12^{\ddagger None,1-2}$
Median quiz score	3.0	$3.0^{\dagger None}$	$7.0^{\ddagger None,1-2}$
Median minutes video watching	21.78	21.87	$102.82^{\ddagger None,1-2}$
Median minutes on platform	94.56	90.83	$542.04^{\ddagger None,1-2}$

almost all metrics (the exception being age) between those learners not (or hardly) using IntelliEye and those using IntelliEye three or more times. The number of learners in each group though—highly skewed with more than 1,600 learners in the not/hardly using IntelliEye groups and 35 learners in the remaining group—has to serve here as a point of caution. Based on these results, IntelliEye appears to be used most often by learners who are already engaged—a finding which is inline with prior MOOC interventions, e.g. [4, 5].

Next, we consider the use of IntelliEye across time (Figure 4): for each day of our experiment we plot the number of learners exposed to IntelliEye and whether they have IntelliEye or non-IntelliEye session. The usage of IntelliEye neither increases nor decreases significantly over time.

In Table 3 we take a look at learners' decisions of enabling or disabling IntelliEye in subsequent video sessions. Learners that enabled IntelliEye in a video session, did so again with a probability of 0.35 (6% of learners chose to enable IntelliEye for all sessions, 29% chose to enable IntelliEye for just the next video session). After enabling IntelliEye in a video session, 21% decided to permanently disable IntelliEye in the next session. We discuss the main reasons for this decision at the end of this section. Learners that disabled IntelliEye in their video session were very unlikely to change their decision in the next video session with 97% of learners sticking to their disable decision.

Next, we consider for *how long* learners are using IntelliEye during their video sessions: do they use IntelliEye continuously or do they disable it after some time? For all the IntelliEye sessions in which IntelliEye is enabled initially (725 sessions from 557 distinct learners), we condense the video session time (which includes video watching as well as other activities on the platform) to video watching time only, based on the edX log data. We then proceed to determine whether IntelliEye was consistently enabled throughout, or whether it was disabled in the first, second or the last third of the video. We find (Table 4) that mostly IntelliEye is either switched off very early or employed throughout a session. Few learners disable it well into the video watching experience (beyond the first third of the video). Learners that received the pause alert are more likely to disable IntelliEye than learners in the other alert groups; learners in the visual alert condition are most likely to keep IntelliEye enabled, reflecting the various levels of disturbance the alerts cause.

Table 3: `IntelliEye` usage transition probabilities between subsequent video sessions; E=Enabled, D=Disabled, EF=Enabled Forever, DF=Disabled Forever.

Decision $v(i)$	Decision $v(i+1)$			
	E	D	EF	DF
IntelliEye enabled	0.29	0.43	0.06	0.21
IntelliEye disabled	0.03	0.68	0.00	0.28

As a last analysis of this research question, we focus on the reasons learners provided when disabling IntelliEye. Of the 938 learners (248 of them have at least one IntelliEye session) who chose to disable IntelliEye forever, 379 provided us with reasons for their decision. With an open card sort we sorted the provided reasons into eight categories shown in Table 5. As the vast majority of learners reported a single reason, for the few (< 10) learners who provided a number of reasons we selected the one they were most vocal about. Most commonly (35%) learners cited themselves as not needing help to self-regulate their learning (*I never lose my attention because the lecture and the whole course are very interesting.*).

22% of the learners mentioned a non-functioning Webcam (e.g. *Because my camera doesn't work well*; *Webcam and audio are easily accessible with WebRTC so I cover and disable it.*), followed by 17% with privacy concerns (e.g. *I feel awkward being observed and controlled.*; *I don't like the idea of having the webcam on.*) and 9% with IntelliEye not performing as expected [9]. Interestingly, conscious multitasking was mentioned several times (*I'm multi-tasking while doing this.*), showing that at least some learners are very much aware of their learning behaviour and what IntelliEye is supposed to do for them. Among the 27 learners who report being disturbed by the alerts, 12 learners received the pause and 12 learners the auditory alert. Overall, this feedback shows that IntelliEye works reasonably well (only 34 out of 248 learners using IntelliEye at least once reported issues) and that the largest issue facing future use of IntelliEye is learners' *perception* of not requiring an attention tracker during their learning, followed by privacy concerns.

Table 4: Number of sessions with `IntelliEye` initially enabled grouped by the time it is switched off in the session.

Disabled during	Pause	Auditory alert	Visual alert
1st third of a session	48%	44%	35%
2nd third of a session	6%	10%	7%
Last third of a session	7%	6%	6%
Enabled throughout	39%	39%	52%
Total # sessions	242	207	276

5.3 RQ3: Impact of IntelliEye

We now investigate the impact of IntelliEye on learners over time and explore whether learners change their video watching behaviour over time. Specifically, we consider all learners with at least two IntelliEye sessions (the most active learner in our dataset has six IntelliEye sessions); for each learner we bin her

[9] We note that one possible reason is our lack of a calibration step: to make IntelliEye easy to use and accessible we did not impose one; IntelliEye assumes the learner to be facing the screen and the Webcam.

Table 5: Reasons provided for disabling IntelliEye forever.

Reason	#Learners	[%]
Attention tracking not perceived as useful/needed	131	35%
Webcam not functioning	83	22%
Privacy concerns	64	17%
IntelliEye not working well	34	9%
Disturbed by alerts	27	7%
Conscious facing away from the screen	14	4%
Hardware/Internet connection too slow	14	4%
Conscious multitasking	6	2%
Uncomfortable feeling	6	2%
$\sum$	379	40%
No reason provided	559	60%

sessions into two bins (the first half and the second half). We then proceed to compute for each bin (i) the average number of minutes lecture videos were played, (ii) the average attention duration and inattention duration detected by IntelliEye, and, (iii) the average number of inattention alerts occurring per minute of video watching. The results are shown in Table 6. Recall that according to the literature, inattention occurs frequently in video watching, though the manner of investigating this (through probes issued at certain times to study participants) [13, 23, 28] does not allow us to draw minute-by-minute conclusions. In contrast, in our work we can now make a statement to this effect: the average number of inattention alerts varies between 0.84 and 2.86 per minute (the latter means that on average a learner gets distracted every 21 seconds in the visual alert condition!). Across all conditions, on average 1.65 inattention alerts are triggered per minute (i.e. one every 36 seconds on average). Interestingly, learners are quickly able to adapt their behaviour towards the offered technology: while the learners in the visual alert type are often alerted (in a manner that is easy to ignore), the learners in the auditory alert conditions receive significantly fewer alerts (cf. row *Mean #inattention per min*); similarly, learners in the pause and auditory alert conditions have significantly shorter inattention spans (cf. row *Mean avg. inattention duration*) than those in the visual condition. As learners were assigned to the conditions randomly we are confident that this behavioural adaptation is due to the different types of alerts.

When comparing the statistics for the two session bins (to detect trends over time), we do not observe a significant decrease over time in the number of inattention triggers per minute and the duration of inattention. There are a number of reasons that can explain this outcome (e.g. as the material becomes more difficult over time, maintaining the same attention levels may already be a success), we will leave this investigation to future work.

6 CONCLUSIONS

In this paper, we have tackled an issue that is inhibiting successful learning in MOOCs: learners' ability to self-regulate their learning. We have designed IntelliEye to increase learners' attention while watching MOOC lecture videos by alerting learners to their loss of attention (approximated through face tracking via Webcam feeds) in real-time. To re-gain learner attention, we trialed three types of interventions—pausing the video with automatic resume once the learner is focusing on the video again, an auditory alert to call learners to attention, and a visual alert around the video widget.

Table 6: Overview of the impact of IntelliEye on learners' behaviors. There are 37 (pause), 27 (auditory) and 41 (visual) learners in each group. † indicates significance at $p < 0.05$ level between the first half and the second half of the IntelliEye sessions (Mann-Whitney U test). ∗ indicates significance at $p < 0.05$ level between the marked group and the visual alert group (Mann-Whitney U test).

Metrics	Alert type	First 50% IntelliEye sessions	Last 50% IntelliEye sessions
Mean avg. video playing length (min)	Pausing	11.93(9.46)	15.96(13.38)∗
	Auditory alert	13.38(10.43)	16.16(13.17)
	Visual alert	17.15(16.21)	24.68(20.38)†
Mean avg. attention duration (min)	Pausing	6.71(7.09)	6.70(8.94)
	Auditory alert	9.38(8.76)	9.04(12.13)
	Visual alert	9.33(9.40)	12.53(17.35)
Mean avg. inattention duration (min)	Pausing	0.62(1.45)∗	0.50(1.25)
	Auditory alert	0.45(1.94)∗	1.07(4.93)∗
	Visual alert	3.69(9.03)	3.46(5.29)
Mean avg. #inattention per min	Pausing	1.30(1.96)	1.50(2.13)
	Auditory alert	0.84(2.05)∗	0.93(2.14)∗
	Visual alert	2.86(4.31)	2.13(3.24)

To explore the viability and acceptance of learners towards such an assistive system, IntelliEye was deployed in an engineering MOOC across a 74-day period to 2, 612 learners.

Our analyses explored three issues: (1) the technological capabilities of our MOOC learners' hardware, (2) the acceptance of IntelliEye by MOOC learners, and, (3) the effect of IntelliEye on MOOC learners' behaviour. We found the vast majority of learners (78%) to possess hardware capable of running IntelliEye; we found fewer—though still a considerable number—learners willing to try such an assistive tool (32% of all learners with supported setups) and among those that did use IntelliEye we determined extremely high levels of inattention, on average 1.65 inattention events per minute (i.e. on avg. inattention arises every 36 seconds).

Learners learnt to adapt their behaviour as needed: learners in the pausing/auditory conditions had significantly fewer inattention events than learners in the non-disruptive visual alert condition. This though, did not yet translate into learning gains. Learners that opted not to use IntelliEye often did not see a need for it and were concerned about their privacy.

Considering the facts that we observe high levels of inattention and that learners once they make a decision on the tool's usage do not change that decision, we need to put more effort into the initial "sign-up" phase of such a tool in future work.

With IntelliEye being the first of its kind to address the learner (in)attention problem in MOOCs in real-time and by relying on non-calibrated common Webcams and open-source face tracking, we have shown that there is a potential for such a system. In our future work, we will extend the deployment of IntelliEye to a larger audience and a wider variety of MOOCs. We will investigate learner incentives and compliance issues to increase the awareness and acceptance of our approach.

ACKNOWLEDGEMENTS

This research has been partially supported by the EU Widening Twinning project TUTORIAL, the Leiden-Delft-Erasmus Centre for Education & Learning and NWO project SearchX (639.022.722).

REFERENCES

[1] Robert Bixler and Sidney D'Mello. 2014. Toward fully automated person-independent detection of mind wandering. In *UMAP'14*. Springer, 37–48.

[2] Robert Bixler and Sidney D'Mello. 2016. Automatic gaze-based user-independent detection of mind wandering during computerized reading. *User Modeling and User-Adapted Interaction* 26, 1 (2016), 33–68.

[3] Diane M Bunce, Elizabeth A Flens, and Kelly Y Neiles. 2010. How long can students pay attention in class? A study of student attention decline using clickers. *Journal of Chemical Education* 87, 12 (2010), 1438–1443.

[4] Dan Davis, Guanliang Chen, Tim Van der Zee, Claudia Hauff, and Geert-Jan Houben. 2016. Retrieval practice and study planning in moocs: Exploring classroom-based self-regulated learning strategies at scale. In *European Conference on Technology Enhanced Learning*. Springer, 57–71.

[5] Dan Davis, Ioana Jivet, René F Kizilcec, Guanliang Chen, Claudia Hauff, and Geert-Jan Houben. 2017. Follow the successful crowd: raising MOOC completion rates through social comparison at scale.. In *LAK*. 454–463.

[6] 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, 41–50.

[7] James Hartley and Alan CameronâĹŮ. 1967. Some observations on the efficiency of lecturing. *Educational Review* 20, 1 (1967), 30–37.

[8] Alex H Johnstone and Frederick Percival. 1976. Attention breaks in lectures. *Education in chemistry* 13, 2 (1976), 49–50.

[9] Katy Jordan. 2014. Initial trends in enrolment and completion of massive open online courses. *The International Review of Research in Open and Distributed Learning* 15, 1 (2014).

[10] 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, 2095–2102.

[11] René F Kizilcec, Mar Pérez-Sanagustín, and Jorge J Maldonado. 2016. Recommending self-regulated learning strategies does not improve performance in a MOOC. In *Proceedings of the Third (2016) ACM Conference on Learning@ Scale*. ACM, 101–104.

[12] Sophie I Lindquist and John P McLean. 2011. Daydreaming and its correlates in an educational environment. *Learning and Individual Differences* 21, 2 (2011), 158–167.

[13] Kep Kee Loh, Benjamin Zhi Hui Tan, and Stephen Wee Hun Lim. 2016. Media multitasking predicts video-recorded lecture learning performance through mind wandering tendencies. *Computers in Human Behavior* 63 (2016), 943–947.

[14] Eduardo Lundgren, Thiago Rocha, Zeno Rocha, Pablo Carvalho, and Maira Bello. 2015. tracking. js: A modern approach for Computer Vision on the web. *Online]. Dosegljivo: https://trackingjs. com/[Dostopano 30. 5. 2016]* (2015).

[15] H Maddox and Elizabeth Hoole. 1975. Performance decrement in the lecture. *Educational Review* 28, 1 (1975), 17–30.

[16] John McLeish. 1968. *The lecture method*. Cambridge Institute of Education.

[17] George A Miller. 1956. The magical number seven, plus or minus two: some limits on our capacity for processing information. *Psychological review* 6a3, 2 (1956), 81.

[18] Caitlin Mills, Robert Bixler, Xinyi Wang, and Sidney K D'Mello. 2016. Automatic Gaze-Based Detection of Mind Wandering during Narrative Film Comprehension. In *EDM'16*. 30–37.

[19] Lloyd Peterson and Margaret Jean Peterson. 1959. Short-term retention of individual verbal items. *Journal of experimental psychology* 58, 3 (1959), 193.

[20] Phuong Pham and Jingtao Wang. 2015. AttentiveLearner: improving mobile MOOC learning via implicit heart rate tracking. In *International Conference on Artificial Intelligence in Education*. Springer, 367–376.

[21] Paul R Pintrich and Elisabeth V De Groot. 1990. Motivational and self-regulated learning components of classroom academic performance. *Journal of educational psychology* 82, 1 (1990), 33.

[22] Keith Rayner. 1998. Eye movements in reading and information processing: 20 years of research. *Psychological bulletin* 124, 3 (1998), 372–422.

[23] Evan F Risko, Nicola Anderson, Amara Sarwal, Megan Engelhardt, and Alan Kingstone. 2012. Everyday attention: variation in mind wandering and memory in a lecture. *Applied Cognitive Psychology* 26, 2 (2012), 234–242.

[24] Tarmo Robal, Yue Zhao, Christoph Lofi, and Claudia Hauff. 2018. Webcam-based Attention Tracking in Online Learning: A Feasibility Study. In *23rd International Conference on Intelligent User Interfaces*. ACM, 189–197.

[25] Kshitij Sharma, Patrick Jermann, and Pierre Dillenbourg. 2015. Displaying Teacher's Gaze in a MOOC: Effects on Students' Video Navigation Patterns. In *Design for Teaching and Learning in a Networked World*. Springer, 325–338.

[26] Jonathan Smallwood, Daniel J Fishman, and Jonathan W Schooler. 2007. Counting the cost of an absent mind: Mind wandering as an underrecognized influence on educational performance. *Psychonomic bulletin & review* 14, 2 (2007), 230–236.

[27] John Stuart and RJD Rutherford. 1978. Medical student concentration during lectures. *The lancet* 312, 8088 (1978), 514–516.

[28] Karl K Szpunar, Novall Y Khan, and Daniel L Schacter. 2013. Interpolated memory tests reduce mind wandering and improve learning of online lectures. *Proceedings of the National Academy of Sciences* 110, 16 (2013), 6313–6317.

[29] Karen Wilson and James H Korn. 2007. Attention during lectures: Beyond ten minutes. *Teaching of Psychology* 34, 2 (2007), 85–89.

[30] Yue Zhao, Christoph Lofi, and Claudia Hauff. 2017. Scalable Mind-Wandering Detection for MOOCs: A Webcam-Based Approach. In *European Conference on Technology Enhanced Learning*. Springer, 330–344.