Glancee: An Adaptable System for Instructors to Grasp Student Learning Status in Synchronous Online Classes

Shuai Ma
The Hong Kong University of Science and Technology
Hong Kong, China
shuai.ma@connect.ust.hk

Taichang Zhou*
Fei Nie*
The Hong Kong University of Science and Technology
Hong Kong, China
tzhouam,fnie@connect.ust.hk

Xiaojuan Ma
The Hong Kong University of Science and Technology
Hong Kong, China
mxj@cse.ust.hk

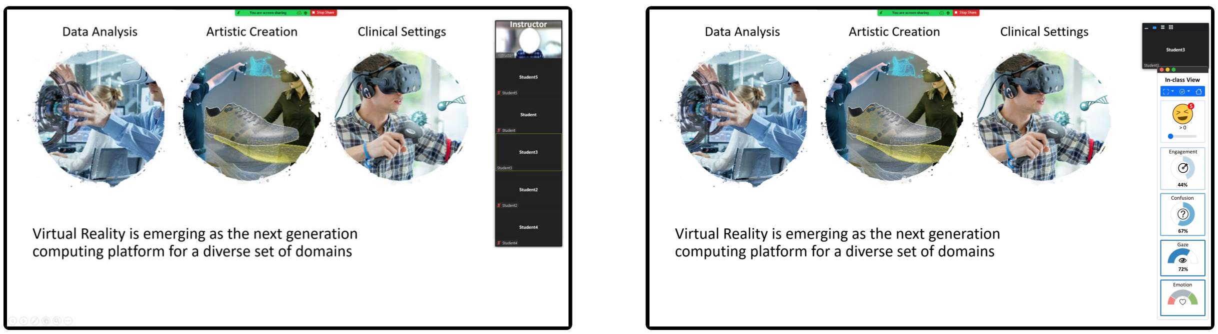


Figure 1: This figure shows a common scene where an instructor is delivering an online class with a video conferencing platform (e.g., Zoom). Usually few students are willing to show their videos, which hinders instructors' ability to read the classroom (left). We propose *Glancee* which conveys students' learning status to instructors via a sidebar interface during the class (right).

ABSTRACT

Synchronous online learning has become a trend in recent years. However, instructors often face the challenge of inferring audiences' reactions and learning status without seeing their faces in video feeds, which prevents instructors from establishing connections with students. To solve this problem, based on a need-finding survey with 67 college instructors, we propose *Glancee*, a real-time interactive system with adaptable configurations, sidebar-based visual displays, and comprehensive learning status detection algorithms. Then, we conduct a within-subject user study in which 18 college instructors deliver lectures online with *Glancee* and two baselines, *EngageClass* and *ZoomOnly*. Results show that *Glancee* can effectively support online teaching and is perceived to be significantly more helpful than the baselines. We further investigate how instructors' emotions, behaviors, attention, cognitive load, and trust are affected during the class. Finally, we offer design recommendations for future online teaching assistant systems.

*Both authors contributed equally to this research.

CHI '22, April 29-May 5, 2022, New Orleans, LA, USA

ACM ISBN 978-1-4503-9157-3/22/04…$15.00
https://doi.org/10.1145/3491102.3517482

CCS CONCEPTS

• **Human-centered computing → Empirical studies in HCI**; • **Applied computing → Computer-assisted instruction**.

KEYWORDS

E-Learning, Online Class, Affective Computing, Videoconferencing, Human-centered Design

ACM Reference Format:
Shuai Ma, Taichang Zhou, Fei Nie, and Xiaojuan Ma. 2022. Glancee: An Adaptable System for Instructors to Grasp Student Learning Status in Synchronous Online Classes. In *CHI Conference on Human Factors in Computing Systems (CHI '22), April 29-May 5, 2022, New Orleans, LA, USA*. ACM, New York, NY, USA, 25 pages. https://doi.org/10.1145/3491102.3517482

1 INTRODUCTION

In recent years, synchronous online classes have become a widespread education model among universities, especially during the COVID-19 pandemic [17, 96, 98]. This model makes it convenient for learners to attend lectures anywhere through remote communication tools, e.g., Zoom [30, 90]. It is reported by UNESCO that by the time of submission of the paper, nearly half of the world's students are still affected by partial or full school closures [88]. Some educators postulate that this online class mode may remain prevalent and even be standardized in the post-pandemic era [91].

Despite its popularity and significance, moving synchronous classes online imposes various challenges on college instructors [16, 17, 98]. Among them, the lack of direct face-to-face interaction is one of the most severe problems [17, 98], as seeing students' faces is a conventional means for instructors to better connect with audiences, master the dynamics of the classroom, and make targeted adjustments to the teaching rhythm and content [17, 79, 98]. For one thing, existing video conferencing platforms often prioritize showing the presenter slides with only a small number of participant video windows, leaving limited communication bandwidth for instructors to monitor students' learning status [54]. For another, many students are not willing to show their videos due to concerns such as privacy [98], and thus it further hinders instructors' abilities to read the classroom and adjust their teachings accordingly.

In offline or asynchronous online teaching scenarios, some teaching assistant systems have been proposed to help instructors grasp students' learning status. For example, EngageMeter [37] helps teachers master students' engagement, EmotionCues [100] analyzes students' emotions, and MudSlide [35] collects content that students get confused about. Recently, Sun et al. [79] proposes similar support for synchronous online teaching, capturing and displaying students' flow-related psychological states in real-time. These prior works demonstrate the potential of inferring students' learning status from their video streams by leveraging advanced computer vision-based techniques. Nevertheless, there remain two key research gaps. First, existing literature has proposed various types of student learning status merely based on traditional educational theories [20, 60, 74, 75]. However, there is a lack of empirical investigation on which type(s) of learning status information instructors would like to see during synchronous online teaching in practice. Second, existing works usually directly present a fixed type of learning status to instructors in a pre-defined design, where instructors have limited control over what and how information could be presented. If indeed instructors have different preferences of students' learning status and how to display the data while teaching, existing tools lack the adaptability to accommodate such needs. And correspondingly, little is known about how instructors perceive and use an adaptable system and how an adaptable system could affect instructors. To mitigate these research gaps, we raise and explore five research questions (RQ1-5).

First, to investigate instructors' actual information needs and preferences and the feasibility to provide such information in synchronous online teaching scenarios, we aim to address two questions: **RQ1: What kinds of students' learning status are instructors concerned about in synchronous online classes?** and **RQ2: What would be a method acceptable for students to capture and display their learning status on the fly?** To answer **RQ1**, based on a literature review, we design and launch a survey with 67 instructors from different majors in a local research university. From the survey results, we find that there are obvious individual differences among the instructors in terms of the perceived importance of various types of student learning status. Besides, we obtain valuable insights into target users' needs and preferences of how to display students' learning status in an interface, confirming the necessity to design an adaptable system (the second research gap). To answer **RQ2**, we survey 62 students from different grades and majors in a local university and get their concerns and acceptance of possible methods to collect and display their learning status data in real-time. The results show that students are willing to accept the vision-based learning status detection method under the premise that their privacy and anonymity are well protected. Taking the informative feedback from both instructors and students into consideration, we obtain five design requirements and organize a design workshop to discuss the potential interface design to cater to instructors' needs of adaptability, space-saving, little distraction and low burden. Based on the outcomes, we propose *Glancee*, an adaptable system which enables instructors to grasp students' learning status with a simple glance. At the backend of *Glancee*, we integrate a set of computer vision-based algorithms to detect a series of students' learning statuses. At the frontend, we provide a sidebar-like in-class view for real-time class state monitoring and a post-class view for retrospective review and analysis. To meet individual preferences, the interface is adaptable, allowing instructors to customize the interface features (e.g., information to display, visualization style, reminding mechanism, etc.) freely.

With the proposed research prototype, we further explore the following research questions: **RQ3: How are the usability and effectiveness of the adaptable system in real online teaching?** and **RQ4: How will instructors interact with and be influenced by the system during a live online class?**. Since the system tries to compensate for the lack of video feed with AI-inferred information, we are interested in investigating **RQ5: How will instructors trust and collaborate with such an AI-empowered system?** To answer these questions, we conduct a within-subjects user study with 18 instructor participants and 53 student participants from different universities, majors, and backgrounds, where instructors deliver three live online lectures in three different, counter-balanced conditions. We compare the proposed *Glancee* with two baseline conditions: one is called *EngageClass* which involves Zoom and a non-adaptable version of our system showing a single measure – students' engagement level, and the other is *ZoomOnly* without any additional students' learning status display. Through mixed-methods analyses, results from in-task, post-task questionnaires, and interviews show that *Glancee* can effectively support online teaching and is perceived to be significantly more helpful than the two baselines from multiple perspectives. Besides, we observe that instructors tend to take frequent glances at our sidebar-like interface to grasp student learning status during the teaching process and they focus on different learning statuses. Furthermore, instructors adjust their lecture delivery based on the displayed student learning status and their own teaching strategy. In addition, *Glancee* leads to instructors' overall positive emotions, and the cognitive and attentional workload it imposes is considered acceptable. Moreover, instructors are inclined to trust the system, and whether the data displayed matches their expectations plays a crucial role in their trust-building. Based on our findings, we offer practical implications for designing online class assistant systems.

Our key contributions include:

- A survey with 67 instructors on their real needs and preferences in synchronous online teaching scenarios and a survey with 62 students to find an appropriate learning status data collection and display approach.
- An online teaching support system, *Glancee*, integrating an instructor-centered, adaptable interface with real-time learner status inference algorithms for instructors to read the classroom and establish connections with students.

- A mixed-methods user study to investigate the usability and effectiveness of the proposed *Glancee*; to explore instructors' interaction patterns and the effects on their behaviors, emotions, attentional and cognitive workload; and to understand instructors' trust and collaboration issues when using the system in actual synchronous online teaching.

2 RELATED WORK

2.1 Context and Challenges in Synchronous Online Classes

The synchronous online class has become a prevalent paradigm adopted almost worldwide ever since the outbreak of COVID-19 [17, 38, 48, 62]. To conduct online teaching, various platforms have been utilized, including social media, videoconferencing, professional teaching platforms, etc [16]. The most popular method is utilizing video conferencing platforms, such as Zoom, Microsoft Teams, Skype, Adobe Connect [30, 90], where instructors deliver a lecture to remote students in real-time with slides on a shared screen. Current video conferencing platforms provide some interaction supports to users, such as audio, video, text box, vote, emoji, breakout rooms, etc [16, 17]. However, there remain a lot of problems and challenges in current video conference-based online classes for instructors and students [16, 17, 98].

One of the most severe problems claimed by instructors is the lack of connection with students, as they find it difficult to perceive students' learning statuses and get timely feedback [16, 17, 98]. For example, according to [17], almost all instructors in their interviews expressed concern about not seeing students' faces during live streaming sessions. For one reason, students always hide their faces if showing video is not compulsory [98]. For example, it is reported in [98] that 80% of students were reluctant to share their video feeds due to reasons such as discomfort and privacy concerns. For another, existing video conferencing platforms often prioritize showing the presenter slides with only a limited number of participant video windows, leaving limited communication bandwidth for instructors to monitor students' learning status [54]. This will hinder instructors' ability to read their classroom and make adjustments to their teaching, which can even make instructors feel like they are "talking into a void" [98]. Similarly, from the students' perspectives, they usually felt struggled to connect with peers and instructors [16, 98] and were unable to get instant responses from instructors [17] due to the lack of face-to-face interaction. Therefore, in this paper, we focus on the problem that instructors can not easily observe instant students' reactions and states due to students' unwillingness to show their videos, and propose a method to help establish a connection between instructors and students.

2.2 Student Learning Status Detection and Display

To get a pre-selected list of common student learning statuses for the exploratory survey design, we did a literature review to collect students' learning behaviors/states detected in existing computer vision-based teaching support systems. Specifically, based on academic databases (e.g., Google Scholar, ACM digital library, IEEE Xplore), we first searched by relevant keywords including "online learning", "student learning states", "learning engagement", "audience feedback", "student sensing", etc. All related journal and conference papers published in the recent ten years were listed for further refinements. After reviewing the contents of these papers, twenty papers were selected as relevant materials. Then we iteratively reviewed the *"cited by"* and *"reference"* of these papers. Finally, we collected a list of states/behaviors from 30 relevant papers and categorized them into five types, namely *state*, *emotion*, *head/facial behavior*, *gaze behavior*, and *upper-body behavior*, as shown in Table 1.

For *state* detection, there are two main kinds of methods. One is based on sensor data analysis, where physiological signals, such as heart rates [61, 93], EEG signals [37, 46], skin temperature [67], electrodermal activity [25] are sensed and rule-based or ML-based methods are applied to predict the states. The other is based on computer vision approaches, where facial features [54, 79], gaze [26, 42], head motion [32, 65] were collected to predict students' affective and cognitive states. For *emotion* detection, a great number of computer vision-based algorithms have been proposed [8, 53] to recognize the emotions from facial images. At the same time, a lot of open-sourced APIs are easy to access, such as dlib [44], Microsoft Face API [5], etc. For *head/facial behaviors* detection, there are still a lot of open-sourced tools, such OpenFace [6], OpenCV [58], etc. For *gaze behaviors* detection, there are two main kinds of approaches. One is using commercial Eye trackers to detect gaze movement, such as Tobii. The other is the computer vision-based method [2, 101, 103] which only relies on a webcam. For *upper-body behaviors* detection, OpenPose [14], Microsoft Kinect [105] have been widely adopted in HCI research.

For the display of the detected student learning status, some commonly used methods, such as chart-based visualization [37, 79], color visualization [32, 85], AR/VR embodied displays [19, 39, 99], raw facial videos [54, 66], videos aligned with visual elements [1], ambient visualization [4] have been adopted in different online teaching systems.

In this paper, we focus on the synchronous online class scenario. Considering the feasibility of the sensing devices and the instructors' requirements to see the overall states of the class, we decided to apply computer vision-based methods to detect students' states and behaviors, and display the information via intuitive chart-based visualization.

2.3 Existing Systems for Audience Sensing and Feedback to Presenter

A wide variety of systems have been proposed to transmit audiences' feedback to stakeholders (the presenter or the audiences themselves) [15, 35, 37, 68, 85, 87, 107]. In the following, to align with our target users (i.e., instructors), we focus on related work providing a group of audiences' feedback to the presenter.

Traditional methods explicitly collect audience feedback via different kinds of self-reports [15, 35, 68, 85]. For example, Interest Meter [68] and iClicker [15] have been proposed to gather responses from the audience during the class and the results were aggregated to instructors. Glassman et al. proposed Mudslide [35] for

Table 1: Student learning status detected in existing teaching support systems.

Category	State/Behavior
State	engagement [23, 26, 33, 37, 46, 78, 82, 94], confusion [35, 54], attention [42, 64, 65], thinking [54], concentration [75], flow [75, 79], mind wandering [41, 42].
Emotion	interest [75], frustration [54], excitement [54], enjoyment [75], anxiety [79], surprise [54, 100], boredom [26, 42, 54, 79], happiness [54, 100], sadness [54, 100], neutral [54, 100]
Head/facial behavior	head-shaking and head-nodding [54], speaking [1, 54], smiling [1, 54], frowning [54], laughing [54], yawning [54], drowsiness [54], head still [65]
Gaze behavior	eye closed [54], looking away [54], gaze fixation and saccade [9, 42], gaze out of screen [42], located on target [2, 9, 77, 97]
Upper-body behavior	raising hand [1, 54], thumbs up [54, 85], clapping [54], leaning backward&forward [54], hand touching face [1, 54], arms at rest [1]

students to anchor confusing points on the slides and help instructors to understand students' confusion. Teevan et al. [85] developed a smartphone-based application for audiences to indicate their thumbs up/down in real-time and visualize the results along with presenters' slides. Although the explicit methods have some strengths, such as raising audiences' engagement [85], there are some shortcomings, including the increased cognitive workload and distraction for both the presenter and audience [68, 85]. In addition, audiences might forget to provide feedback due to some reason, such as being too engaged or disengaged [54].

In the meantime, a number of implicit methods have been proposed to collect audience feedback, such as EngageMeter [37], AttentivU [46], Sync Class [32], EduSense [1], EmotionCues [100], etc. However, these works focused on the offline in-person classroom and required high-resolution camera groups or physiological sensing devices, which is infeasible to be applied in the distributed online learning setting. For synchronous online presentations, [54] proposed a method to "spotlight" the most expressive audience's video feed to make presenters more aware of their audience. However, this method assumes the audience's use of video, but most students are unwilling to show their videos in an online class [16, 17, 98]. The most similar work to this paper might be [79]. They developed a system that monitored the facial expressions of students to classify students' states into one of the three categories (anxiety, flow, boredom) and visualize them to instructors. This approach can deal with students' unwillingness to show videos but still faces several problems. First, it provides a fixed measure (i.e., flow) that may not meet the needs of different teachers. Second, it utilizes an additional screen to display students' states, which is inconvenient for instructors who only have a single monitor with limited screen space. Third, they do not systematically measure the usability and effectiveness of their design. In this work, we adopted an instructor-centered design considering the individual preference and the space constraints of the screen. Through a mixed-methods user study, we systematically explored the usability and effectiveness of the system, and comprehensively investigated the effects on instructors.

3 EXPLORATORY SURVEY

3.1 Methods

We conducted two exploratory surveys to explore RQ1 and RQ2. To explore **RQ1**, we designed a survey to investigate instructors' actual needs in synchronous online classes. First, based on the aforementioned collected student learning status (shown in Table 1), we designed questions to acquire instructors' perceived importance of each learning state/behavior on a 3-point scale, *Very important*, *Somewhat important*, and *Not important*. We specified the *gaze located on target* to *gaze following slides* for our scenario. In the survey, we also asked for instructors' demographic information and teaching experience, how they perceive the situation that students do not show videos in online classes, and whether they use an additional monitor to give online classes. In addition, we asked about their preference of the screen space allocation if the students' learning status information was displayed on their screen. Finally, we requested them to give further comments or suggestions about designing an online class support system.

To explore **RQ2**, we designed a survey for students. The survey includes demographic and background questions, their experience in taking synchronous online classes, and their perceptions of turning on cameras. Then, we requested their concerns and suggestions on building a system to capture their learning status on the fly. After obtaining the institutional IRB approval, we launched the two surveys in a local university via e-mails and social media posts.

3.2 Results

Following [54], we analyzed the quantitative data with descriptive statistics, and two authors coded the responses of open-ended questions using thematic analysis methods [13] and discussed together.

3.2.1 Survey of instructors. We received a total of 67 responses from instructors (44 male, 18 female, four non-binary, and one prefer not to say), with an average age of 46.4 (*SD*=12.0). They came from different schools, such as engineering, science, business, humanity, and they had experiences of teaching online classes for undergraduates and graduates. All participants had at least one semester's online teaching experience due to the COVID-19 pandemic (15.5% one semester, 22.5% two semesters, 39.4% three semesters, 14.1% four semesters, and 8.5% more than four semesters). 83% of instructors had faced situations in which only a few students' cameras were turned on. Among them, 76% of instructors thought it lead to trouble for online teaching. Further, we found that 44% of them did not give online courses with multiple monitors, indicating that displaying student status on an additional screen is not applicable for all teachers. For each kind of student learning states/behaviors, instructors rated with *Very important*, *Somewhat important*, or *Not important*. We ranked all the learning status by a simple normalized

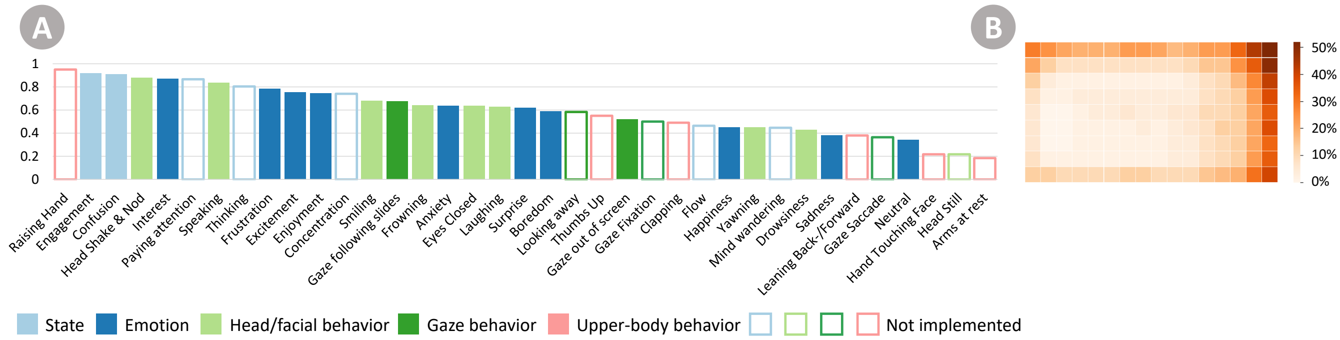


Figure 2: Instructors' feedback on their preferences of student learning status and screen space allocation. (a) The ranking of instructors' perceived importance levels of students' learning status based on a normalized weighted importance score (from 0 to 1, the closer to 1, the more important). The information without color filling represents the student status that is not implemented in the system. (b) Instructors' preference of where to display the students' learning status on a 16:9 screen.

weighted importance score:

$$Score = \frac{\#Very\ important \times 2 + \#Somewhat\ important \times 1}{\#All\ ratings \times 2} \quad (1)$$

Figure 2 (a) shows the importance score (from 0 to 1) of each factor in a descending order. We found individual preference exists because even for the top-ranked learning status, some teachers still perceive it unimportant, and vice versa (original importance ratings can be found in Figure 11, Appendix). In the experiment, we will explore teachers' preferences in more detail. In this paper, we implemented most of these factors (shown in colored columns). However, due to redundancy and other reasons, we did not implement some of them (detailed in Sec 4.2). Figure 2 (b) shows the heatmap of instructors' preferred regions of the screen to display students' learning status, where we can find that most instructors preferred the right side of the screen (1/8 of the screen width).

Finally, we got some valuable feedback from instructors, which can be concluded as follows. First, instructors want more control over the interface, such as "*which information to display*" and "*where the information is displayed*". Second, the display of information should not occupy too much space. Third, the presentation of information should be as intuitive and straightforward as possible to avoid introducing too much burden on instructors (also suggested by [79]).

3.2.2 Survey of students. We received a total of 62 responses from students (30 male, 30 female, 2 prefer not to say) with an average age of 20.3 (*SD*=2.0). They came from different grades and majors, including computer science, business, finance, civil engineering, electronic engineering, mechanical engineering, physics, math, biology, etc. All students had experienced synchronous online classes (17.7% for two semesters, 58.1% for three semesters, 19.4% for four semesters, 4.8% for more than four semesters). From their feedback, only 8.1% of students would like to turn on cameras in online classes, which is consistent with results in previous study [98]. Most students will not turn on the camera unless it is required by the instructor (54.8%), or other classmates have turned on the camera (21.0%). When asked about "*If your video is only used to extract some learning states to display to instructors, and your video will not be shown to others and stored, would you mind turning on the camera?*", 71.0% of students do not mind turning on cameras if their "*image data is protected and not shown*". From the open-ended question, we found that the main reason why the remaining 29.0% of students still mind turning on the camera is that they do not want the instructor to see their personal learning states. For example, some students mentioned "*worrying that the teacher will give participation marks based on students' learning states*". This result reveals the importance of the anonymity of the learning status data. In addition, students offered some suggestions. First, only basic states should be detected. For example, some students commented "*It should not capture private data, such as the content on my screen and the things in my room.*" Second, the video data must be protected. For example, some students commented "*Videos should be processed on my own laptop and not be uploaded to anywhere.*".

To conclude, we can get five design requirements from the two surveys. **R1**: Students' image/video data should be well protected. **R2**: Students' learning status should be anonymously shown to the instructors. **R3**: Instructors should have control over the interface to adapt to their different preferences. **R4**: The interface should not occupy too many areas of the screen, and the display should be applicable for a single monitor. **R5**: The interface should not distract the instructors and should not bring too much burden on instructors when they are delivering courses.

4 DESIGN OF GLANCEE

4.1 System Overview and Architecture

Based on the design requirements derived from our surveys, we propose *Glancee*, an adaptable, interactive system for instructors to observe students' learning status during synchronous online classes. *Glancee* contains three components (Figure 3): a student end, a server, and a web-based teacher end. 1) The student end captures students' images via a webcam and detects learning status by computer vision-based algorithms. To protect students' data (**R1**), the student end runs on their local devices. Once processed, the images will not be saved anywhere. 2) The server receives anonymous data from distributed student ends (**R2**) and aggregates individual data to compute the overall class-level states, i.e., the average level of engagement, the sum of the occurrence of each kind of head/facial behaviors and emotions, the percentage of confusion and gaze. 3) The teacher end gets the class-level states from the server and visualizes them to instructors. In this paper, we focus on conveying students' learning status to instructors instead of building a new video conferencing platform, so we use Zoom as the basic online class platform, and our system provides an assistant interface. Note that our system can be easily integrated with any video conferencing platform, not limited to Zoom.

4.2 Student End: Learning Status Detection

Based on the importance ranking of different learning statuses shown in Figure 2 (a), we selected a set of learning statuses to implement and omitted the rest according to the following criteria: 1) Feasibility of existing computer vision-based algorithms: we did not implement *thinking* because it is hard to achieve by existing vision-based techniques; 2) Redundancy of the information: (a) most *upper-body behaviors* have already been supported in Zoom and are less likely to be performed physically by students in the online class scenario, such as *raising hand*, *thumbs up* and *clapping*, so we did not repeatedly implement this category; (b) *paying attention*, *concentration*, *mind wandering*, *flow* are overlapped with *engagement*, so we only implemented the one with the highest importance score; (c) *head shaking/nodding* and *head still* are complementary, so we chose the more important one; 3) Representativeness: we selected *gaze following slides* and *gaze out of screen* as two representative behaviors in the gaze behaviors category, and ignored *looking away*, *gaze fixation* and *gaze saccade* because these three behaviors are too specific and cannot provide users much high-level information. Next, we introduce the detection algorithms for the selected learning status.

1. State. For students' state, we focus on two important factors, engagement and confusion.

Engagement can be defined as humans' participation and involvement in an interaction, including emotional engagement, behavioral engagement, and cognitive engagement [76]. To train an engagement detection model, we used the "Engagement in the wild" dataset [24, 43], in which students were watching five-minute-long MOOC videos in a diverse environment (more details can be seen in [24]). Our task, however, is different from that in [24] as we need to predict engagement in a real-time manner. Suggested by [37, 46, 83], we divided the 5-minute training video into a set of 5-second, non-overlapping segments. Then two authors with online teaching experience (three years on average, including one semester of online teaching) carefully selected the segments in which the student's engagement level was inconsistent with their parental video's label, and re-labeled them separately. After discussion and relabeling, the labels from the two authors reached a Cohen's kappa score of 0.83, and the two authors resolved conflicts together. Next, following [86], we utilized OpenFace [6] to extract the facial features including *Facial Action Units* related features (AUs) [29], *Eye Gaze* related features, and *Head Pose* related features (more details about features can be seen in [86]). Then, considering the robustness and efficiency in computing, we trained a Random Forest Regressor model with 200 estimators/trees and achieved a 0.05 MSE score on the validation set (comparable with the SOTA models [92, 106]).

Confusion is an essential state in students' learning process [35]. Suggested by [69, 95], the brow furrowing expression (Action Unit 4 in the Facial Action Coding System (FACS) taxonomy [29]) is a strong indicator of the confusion states, and thus we use OpenFace [6] to directly detect the occurrences of this expression to predict confusion [54].

2. Emotion. We applied the Circumplex model of emotions [71] that presents emotional states as points distributed in a two-dimensional, valence-arousal circular space. The horizontal axis represents the valence (positive or negative) of emotion and the vertical axis denotes the arousal (high or low) of emotion, and the center point represents a neutral emotion [70]. We used a SOTA model proposed in [21] which utilized ResNet50 to extract facial features and MLP stacked on top of the final ResNet50 Conv layer to predict the valence and arousal of students' emotions. Finally, we mapped the predicted valence and arousal values on the 2D space (an emotion wheel).

3. Head/Facial Behavior. For this type of learning status, we extracted *head shaking/nodding*, *speaking*, *smiling/laughing*, *eye closed/drowsiness*, *frowning*, and *yawning*. Generally, we first used OpenFace to detect the 68 facial landmarks [45] and head orientations, and then adopted different ML-based methods or heuristic-based methods to detect each behavior. 1) For *head shaking/nodding*, following [54], we used a Hidden Markov Model (HMM) [84] to calculate the probabilities of the head nod and head shake gestures. In particular, the HMM used the head yaw rotation value to detect head shakes, and the head Y-position of the facial landmarks to detect head nods over time. 2) For *speaking*, we first compute the Mouth Aspect Ratio [3] (MAR, the ratio of distance between the vertical and horizontal mouth landmarks), and then calculate the MAR difference between two adjacent frames in a given interval and compare the number of MAR changes with a certain threshold. 3) For *smiling/laughing*, we applied the detector provided by OpenCV[1] to directly detect the smiling and laughing. 4) For *eye closed/drowsiness*, following commonly used approaches [3, 22], we first detected the six main landmarks around the eye and then calculated the average Eye Aspect Ratio (EAR) of both eyes (the same calculation with MAR) and compared it with a threshold. We set a duration threshold to distinguish it from *drowsiness.* 5) For *frowning*, we used the same prediction method with *confusion.* 6) For *yawning*, we used Mouth Aspect Ratio (MAR), same as *speaking* detection but has a significantly larger threshold. Note that all the thresholds were suggested by related papers or tested with varying facial videos carefully.

4. Gaze behavior. We adopted a recently developed approach [101, 102] which is a convolutional neural network based on the AlexNet architecture [47]. The algorithm can estimate the on-screen location of users' gaze only based on the webcam with minimum calibration. Once obtaining the estimated on-screen gaze location, we can measure *Gaze following slides* and *Gaze out of screen* by judging if the gaze point is inside the slides area and screen boundary. We used win32gui[2] to get the position and size of

[1]https://github.com/opencv/opencv/blob/master/data/haarcascades/haarcascade_smile.xml
[2]https://pypi.org/project/win32gui/

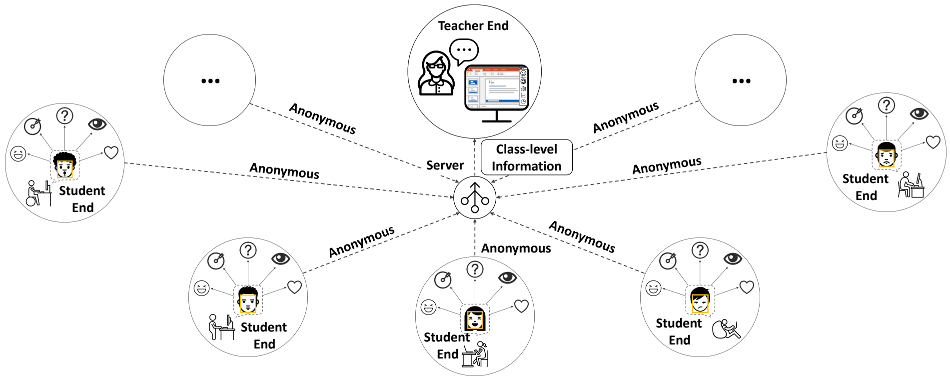


Figure 3: The architecture of the system, including the student end, the server, and the teacher end.

Zoom's window (as an approximation of the slides area). Note that if the Zoom's window is minimized or covered by other windows, the student's gaze is regarded as *not following the slides*.

Finally, we integrated all the learning status detection algorithms into an overall detection module. To balance the efficiency and performance of the algorithms, we set the video image sampling rate to 5 FPS.

4.3 Teacher End: Design of Adaptable Interface for Instructor

4.3.1 Design Workshop. To come up with an appropriate design of the visual display of students' learning status, we organized a workshop with four participants with visualization and HCI background and online teaching experience (2 Females, average age 26). We first presented the design requirements obtained from Sec.3.2. Next, we showed the collected learning status and the corresponding data type, and provided demo data to help designers understand the data comprehensively. After they got familiar with our design objectives and requirements, we conducted a free design session lasting about 20 minutes. Then, each designer presented his/her design alternatives, and other designers gave suggestions and comments. Finally, all participants brainstormed together and proposed the following features.

1) Adaptability. We provide adaptability support to instructors, including *what information, how and where to display* (R3), *whether to turn on the reminder and what remind mechanism to use* to avoid distraction (R5). **2) Sidebar-based In-class view.** Based on instructors' preferences of the screen area (Figure 2 (b)), we display students' learning status in a sidebar form to save space (R4). And the sidebar's size and position can be adjusted flexibly to avoid covering the contents in the slides. **3) Simple Visualization.** Instructors might only have limited attention and cognition resources to perceive extra information(R5), so we present students' learning status with basic charts to make it easier to understand and grasp the information. **4) Separated Views.** We decide to display each type of information in separate views rather than in a single assembled view, because it will cause much burden for instructors to distill certain information from a combined complex chart (R5). **5) Trigger View.** To save space (R4), the head/facial behaviors could be displayed only when triggered because these behaviors do not always exist during students' learning process.

4.3.2 Interface of Glancee*.* The interface of *Glancee* includes three views, a dashboard view, an in-class view, and a post-class view shown in Figure 4.

1. The dashboard view provides flexible configurations of the interface (shown in Figure 4 (a)). 1) *Customization of Visual display.* Users can customize the visual forms for information display in the in-class view and post-class view. For in-class view, users can select radial chart, gauge chart, or bar chart to visualize *engagement*, *confusion*, and *gaze concentration*. And for *emotion* display, users are free to choose charts with different complexity levels (three, five, and nine categories of emotions). For post-class view, users can also specify the chart type (e.g., line chart, area chart) for visual display. In addition, for emotion analysis, we provide a detailed emotion wheel with an overlay heatmap showing the overall emotion distribution during the class. 2) *Customization of Reminder.* Users can decide whether to turn on the notification and which modality (i.e., color highlighting and sound alert) to use. Also, we provide two reminder mechanisms: one is *always remind* which stays alarming as long as the students' states are in poor conditions, and the other is *remind once* in which the reminder is only triggered once in a certain duration. Each learning status has its own triggering condition, and we provide the default thresholds while also allowing users to customize the thresholds individually. 3) *Customization of Color.* Users can change the chart and border color of each view.

2. The in-class view displays student learning status in real-time in a sidebar form (Figure 4 (b)). A total of five separated views are available from top to bottom, including *head/facial behaviors*, *engagement*, *confusion*, *gaze concentration*, and *emotion*. For head/facial behavior, we designed the corresponding facial emoji for each behavior, and the emoji with the largest number of people will be displayed by default. There is a red bubble with a number in the upper right corner of the emoji indicating how many students are showing this behavior. Besides, the in-class view provides 1) *Customization of sidebar size and position.* Users can drag the interface to anywhere on the screen and adjust its size easily by dragging the border. 2) *Control of which view to see.* Users can show or hide any view in the drop-down menu to keep only the needed information, which could reduce the cognitive load. 3) *Change chart type.* During

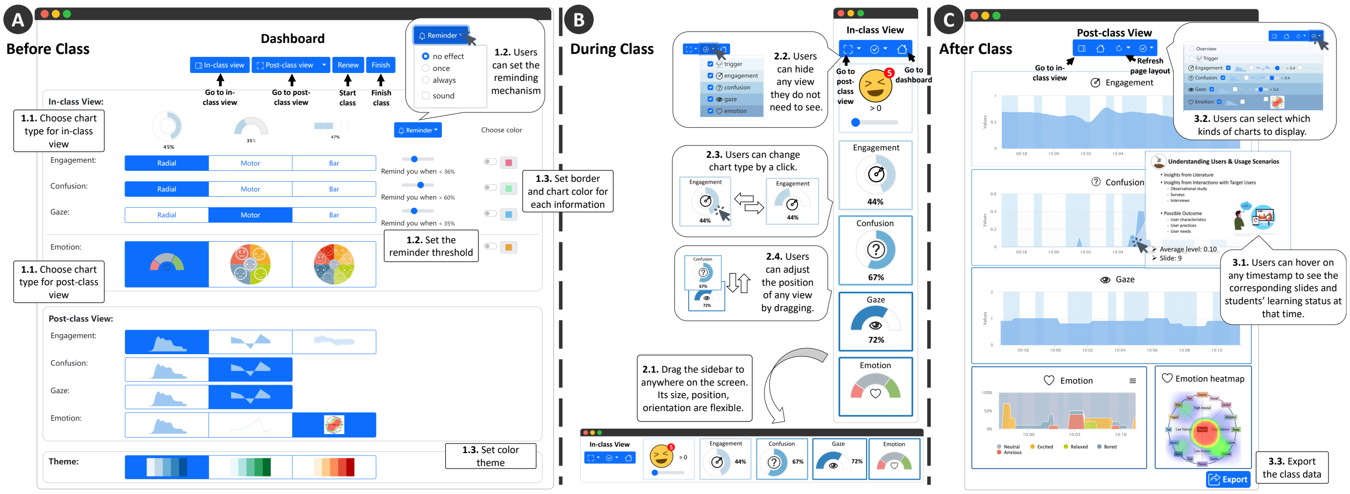


Figure 4: Interface of *Glancee*. (a) The dashboard view for interface customization before class. (b) The in-class view displaying students' real-time learning status during class. (c) The post-class view for reviewing the teaching process after class.

the class, users can quickly change the chart type (e.g., from the radial chart to gauge chart) by a click on the chart. 4) *Reorder the position of each view.* Users can conveniently re-arrange the views shown in the sidebar via dragging. In addition, we utilize a simple tool[3] to avoid the in-class view being blocked by full-screen slides.

3. The post-class view is mainly designed for instructors' post-class review of the teaching process (Figure 4 (c)). We recorded the dynamics of class states throughout a class and plotted them for teachers' inspection. The post-class view also provides some configurable features: 1) *Alignment between learning status chart and slides.* Instructors often want to know how students' learning status was when learning a specific slide [37], thus we align the slides number with timestamps on historical charts. Users can quickly locate the corresponding slide (to see both page number and thumbnail) by hovering the mouse over any position on a curve. To enable this feature, instructors need to upload their slides to our system before the class, and the page number is recognized through the image similarity algorithm based on RGB histogram [81]. 2) *Customization of interface.* Users can select the type of chart to be displayed, adjust the size and position of any chart, so as to create a post-class review interface that better meets their needs and viewing habits. 3) *Export historical data.* Users can export historical records of classroom states into a file for future review or analysis.

4.3.3 Usage Scenario. Mary is an instructor of an HCI course. She gave an online class via Zoom using *Glancee*. Before the class, she first configured the interface based on her preferences. Once the class began, Mary opened the in-class view and placed it on the right side of her screen. During the class, Mary viewed students' learning status in real-time and adjusted her course delivery accordingly. For example, when she found students' engagement dropped, she told a joke to attract their attention. After the class, Mary hit the *Post-class View* button to review how the students' learning status changed over time. She found that students' confusion raised to a high level when she was teaching slide 9, so she navigated to that page and found that one design theory was not well explained. Mary added some clarifications and examples to her slides and planned to explain this design theory again in the next class.

[3]https://deskpins.en.uptodown.com/windows

5 USER STUDY

To investigate the research questions **RQ3-RQ5**, we conducted a controlled within-subject user study, where the instructors delivered online classes with the proposed *Glancee* and two other baseline systems as control conditions.

5.1 Conditions

We compared the proposed *Glancee* with two baselines, *EngageClass* and *ZoomOnly*. 1) *EngageClass* is a simplified, non-adaptive version of *Glancee* which only provides students' engagement level to instructors, and there is also an in-class view provided during the class and a post-class view provided after the class. We set this baseline because most existing systems only provide engagement for instructors [23, 33, 37, 46, 94]. 2) *ZoomOnly* is a pure Zoom without any extra support interface. In the study, instructors were asked to give three lectures with different feedback support systems in counterbalanced order (Latin Square) to minimize ordering effects.

5.2 Task and Procedure

In this study, instructors were asked to deliver three lecture sessions with two breaks in between. The slides need to be prepared carefully in advance while satisfying the following criteria: 1) The instructor must be familiar with the content to ensure the quality of the class; 2) It should contain both difficult parts and interesting parts to cover diverse situations; 3) The three lectures must be comparable in difficulty and interest, and should be under the same topic to avoid potential biases. We asked the instructors to send us their slides at least one day in advance so that we could check whether the content met the above criteria. In the meantime, we sent the student end of *Glancee* to the students one day in advance and asked them to run and test the system. To guarantee the success

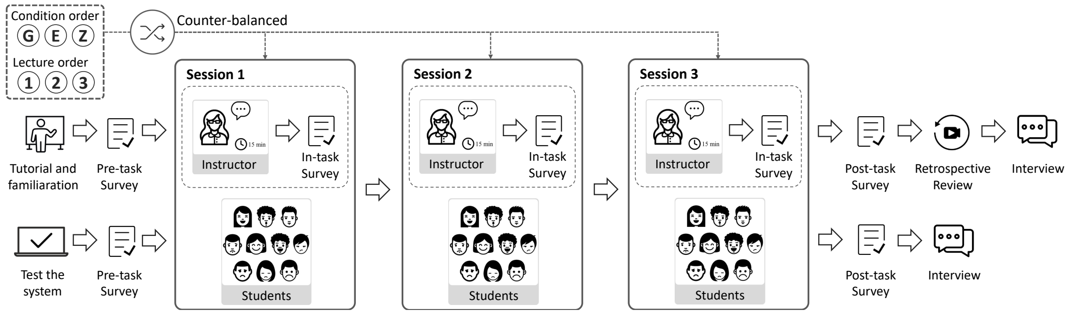


Figure 5: Procedure of the user study. We conducted a within-subject study where each instructor participant needs to give three 15-minute lectures with different systems (G: *Glancee,* E: *EngageClass,* Z: *ZoomOnly*) in a counter-balanced order.

of the experiment, two authors served as experimenters to handle any sudden technical problems and offer the necessary help.

Figure 5 shows the procedure of the experiment. After obtaining participants' consent forms, we first carefully introduced the system to instructors and provided a demo system with randomly generated data for them to get familiar with. Then, we gave instructors and students a pre-task survey to collect their demographics and background information. During the study, with instructors' consent, we kept a video recording of their teaching process and their screen. However, only the slides window needed to be shared with students, and the *student end* kept running across all three sessions so that the experimental conditions were unknown to students to avoid introducing extra variables. During the class, the choice for students to show their videos in Zoom was discretionary. Instructors gave each lecture session via a randomly assigned condition (*Glancee*, *EngageClass*, or *ZoomOnly*). After each session, instructors were asked to recall the teaching process and fill out an in-task questionnaire. After all three sessions, teachers and students filled out a post-task questionnaire. Finally, we reviewed the recordings together with instructors and conducted a semi-structured interview. And we also conducted a semi-structured interview with students. The entire study lasted about two hours for instructor participants (excluding the preparation of slides) and one hour for student participants. Instructors and students received compensations of USD 30 and USD 8 each.

5.3 Participants

After obtaining institutional IRB approval, we recruited both instructor participants and student participants via e-mails and social media posts to have online classes together. For the instructor role, we invited 18 participants (8 Females) with an average age of 29.6 (*SD*=6.7), including lecturers, researchers, assistant/associate professors, and Ph.D. students with Teaching Assistant (TA) experience. The detailed information of teacher participants is shown in Table 2. For the student role, we recruited 53 participants (21 female, 28 male, and 4 prefer not to say) from different universities with an average age of 22.3 (*SD*=1.8), including both undergraduates and graduates. Most of them had online class experience. They came from diverse majors, including computer science, math, electronic engineering, civil, chemistry, biology, mechanical engineering, humanity, etc. Some of them participated in several classes. On average, there were eight audiences in each class, which is a comparable class size with [54, 79].

5.4 Measurement

For instructor participants, we designed a 7-point Likert (1: Not at all/Strongly disagree, 7: Very much/Strongly agree) in-task questionnaire to collect their experience about teaching synchronous online classes with different feedback systems. The in-task questionnaire contained four parts. First, in terms of the **usability** of the systems, referring to [54, 55], we designed questions including 1) ease-of-use, 2) helpfulness in supporting lecture delivery, 3) distraction caused by the system, 4) system satisfaction, 5) likelihood for future use. Second, in terms of the **effectiveness**, referring to [11, 54, 59], we designed questions including 1) facilitating reflection, 2) making the lecture easy to recall, 3) establishing connections with students, 4) raising awareness of the performance of the lecture, 5) enhancing easiness to observe students' reactions, 6) enabling easy adjustment based on students' reactions, 7) ensuring control of the class, 8) perceived lecture satisfaction, 9) perceived attractiveness of the lecture, 10) perceived quality of the lecture. Third, in terms of the impacts on instructors' **emotions**, referring to [28, 31, 52, 54, 79], we designed questions including instructors' feeling 1) engaged, 2) excited, 3) exhausted, 4) frustrated, 5) happy, 6) hopeful, 7) overwhelmed, 8) safe, 9) nervous, 10) anxious, 11) confident. Finally, in terms of the effects on instructors' **workload**, referring to the NASA-TLX survey [36], we designed questions including effects on instructors' 1) cognitive load, 2) attentional load, 3) workload, 4) time spent on students. We also designed a post-task questionnaire to obtain 1) their perceptions of the usefulness of the provided system features, 2) their perceived accuracy of the system and overall trust in the system, 3) their trust in each displayed student learning status. Besides, we collected instructors' interaction logs on the interface. Moreover, we counted the frequency of teachers' glancing at the system. Finally, we conducted a semi-structured interview with teachers asking some open-ended questions. The detailed in-task, post-task questionnaires, and open-ended questions in the interview can be seen in Table 5, Table 6 and Table 7 in Appendix.

For student participants, we designed a post-task questionnaire to explore students' feelings and perceptions during the class, including their feeling 1) nervous, 2) anxious, 3) distracted by the system, 4) feeling, 5) engaged, and 6) future use. Moreover, we

Table 2: Participants information (F = Female M = Male; For the Experience (Exp), A/B, A means Years of Online Teaching Experience, B means Years of Teaching Experience; TA = Teaching Assistant; Assoc.Prof. = Associate Professor; Asst.Prof. = Assistant Professor)

ID	Gender/Age	Exp	Role	Major	Lecture topic
T1	F/27	1/2	TA	HCI	Misinformation in Visualization
T2	M/29	1/3	TA	HCI	Introduction to Tangible Interaction
T3	M/40	2/13	Assoc.Prof.	HCI	Gesture Recognition and Interaction
T4	F/24	1/1	TA	Health Informatics	Cognitive Impairment and Its Rehabilitation
T5	M/23	1/1	TA	HCI	Storytelling and Transitions with Data Videos
T6	F/33	2/6	Lecturer	Mechanical Engineering	Introduction to Axle Design
T7	M/30	0/4	Researcher	HCI	Interaction Design with Non-Verbal Metaphors
T8	M/24	1/1	TA	HCI	Introduction to Social Computing
T9	F/27	1/5	Lecturer	Software Engineering	Interface Design and Development
T10	M/26	1/4	TA	HCI	Prototyping for VR Interaction
T11	F/48	1/27	Lecturer	English	Common Errors Analysis in English Writing
T12	M/24	1/3	TA	Chemistry	Carbon Neutrality and Green Chemistry
T13	M/24	1/1	TA	HCI	Health-related Rumor on Social Media
T14	M/32	1/4	Asst.Prof.	NLP, HCI	Basic Graphics Generation Algorithm
T15	F/32	1/6	Lecturer	Automation	Getting Started with 3D Modeling
T16	F/37	1/12	Assoc.Prof.	Materials Science	Test Methods for Hardness of Materials
T17	M/31	1/5	Lecturer	Mechanical Engineering	Basic Knowledge of Cartography
T18	F/23	1/2	TA	HCI	Automated Suggestions for Non-expert Users

asked open-ended questions to get students' further comments in the interview session. The detailed questions in the post-task questionnaire and interview can be seen in Table 8 and Table 9 in Appendix. Since this part is not our main contribution, we put the questionnaire and interview results of students in Appendix B.

6 RESULTS

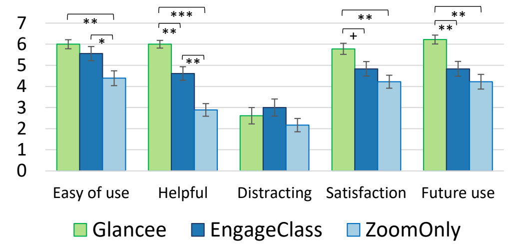
This section presents both quantitative and qualitative results, aiming to answer research questions RQ3-RQ5. For *quantitative* analysis, we ran Friedman Test with post-hoc Wilcoxon signed rank tests on participants' answers in the in-task questionnaires. And we used descriptive statistics to summarize the responses from the post-task questionnaires. For *qualitative* analysis, we followed the steps below. First, once the experiment was completed, we reviewed the screen recording of the lecture session in *Glancee* condition together with the instructors. In the co-reviewing, we confirmed the timestamps of their glancing at the sidebar and asked about the motivations and reactions behind each glancing and each interaction behavior. All feedback was audio-recorded and transcribed into texts by the first author. Since it was a retrospective review process, they were not required to recall all details if they felt vague. Second, two authors coded the co-reviewing transcription and the interview transcription via thematic analysis [40]. These themes were discussed and harmonized over several iterations, and specific examples were identified from the source interviews for use in this paper. In the rest of this section, based on the proposed research questions, we present the results accordingly.

6.1 RQ3: How are the usability and effectiveness of the adaptable system in real online teaching?

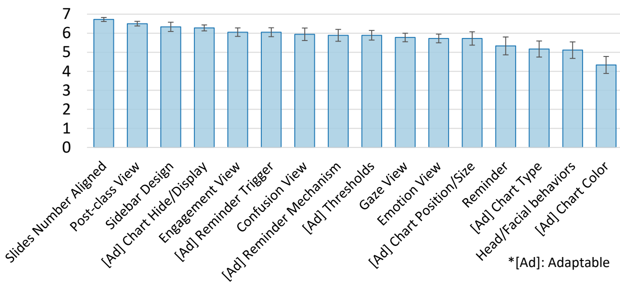
To begin with, we investigate the usability of the proposed *Glancee* perceived by instructors compared with the two baseline systems. Figure 6 (a) shows teachers' average ratings for the usability-related questions in the in-task questionnaire. Friedman tests show that there are significant differences when delivering synchronous online classes with different systems in terms of ease-of-use ($\chi^2(2)$=8.52, p<.05), helpfulness in supporting lecture delivery ($\chi^2(2)$=25.81, p<.001), and likelihood for future use ($\chi^2(2)$=11.82, p<.01). From Post-hoc analysis, *Glancee* is significantly easier to use for inferring students' status compared with *ZoomOnly* (Z=-3.01, p<.01), but shows no significant difference compared to *EngageClass* on this measure. Besides, *Glancee* is significantly more helpful for delivering online classes than both *EngageClass* and *ZoomOnly* (Z=-2.97, p<.01; Z=-3.63, p<.001). *Glancee* was also rated significantly higher than *EngageClass* and *ZoomOnly* in terms of intended future use (Z=-2.77, p<.01; Z=-3.19, p<.01). However, Friedman tests show no significant differences across the three systems in terms of distraction caused by the system and system satisfaction.

Figure 6 (b) shows users' perceived usefulness of different designs and features provided in *Glancee* in descending order. Overall, most features provided in the system were perceived to be useful. Users rated the post-class view highly, especially for the alignment with slide numbers. In addition, users appreciated the design of our sidebar and felt the five types of learning status displayed helpful. At the same time, users found most of the adaptable features useful.

Then, we dig deeper into the whether and how the proposed *Glancee* supported instructors in the synchronous online classes from different aspects. Figure 7 shows instructors' average ratings on the related questions. Friedman tests show that the three systems demonstrate significantly different efficacy in terms of facilitating reflection ($\chi^2(2)$=23.91, p<.001), making lectures easy to recall ($\chi^2(2)$=24.82, p<.001), establishing connections with students ($\chi^2(2)$=23.90, p<.001), raising awareness of the performance of the lecture ($\chi^2(2)$=24.49, p<.001), enhancing easiness to observe students' reactions ($\chi^2(2)$=32.55, p<.001), enabling easy adjustment



(a)



(b)

Figure 6: The usability of usefulness of systems. (a) The usability of different systems perceived by teachers. (b) The perceived usefulness of different features provided by *Glancee*. The error bars indicate standard errors. (+: p<.1; *: p<.05; **: p<.01; ***: p<.001)

based on students' reactions ($\chi^2(2)$=28.96, p<.001), ensuring control of the class ($\chi^2(2)$=27.79, p<.001), as well as perceived lecture satisfaction ($\chi^2(2)$=13.32, p<.01), attractiveness of the lecture ($\chi^2(2)$=12.23, p<.01), and quality of the lecture ($\chi^2(2)$=8.65, p<.05). Post-hoc tests further show that *Glancee* is considered significantly more helpful than *EngageClass* in terms of reflection (Z=-3.41, p<.001), recall of lecture delivery (Z=-3.20, p<.01), connection with students (Z=-2.93, p<.01), awareness of lecture performance (Z=-3.37, p<.001), observation of students' reactions (Z=-3.54, p<.001), adjustment to students' reactions (Z=-2.95, p<.01), control of the class (Z=-3.28, p<.01), and quality of the lecture (Z=-1.98, p<.05). However, no significant difference can be found between *Glancee* and *EngageClass* in terms of lecture satisfaction and attractiveness. When compared with *ZoomOnly*, post-hoc tests show that *Glancee* significantly outperforms this baseline on all measures (p<.01 for satisfaction, attractiveness, and quality; p<.001 for all the rest).

These results are consistent with participants' qualitative feedback in interviews. The reasons why they perceived *Glancee* to be more usable and helpful than the two baselines are that our system provided "*comprehensive student learning status*" which can "*cover most aspects of a class's states*" so it is beneficial for "*reflection and adjustment accordingly*". At the same time, the interface design is "*easy to observe and understand*" without "*causing distraction*".

6.2 RQ4: How will instructors interact with and be influenced by the system during a live online class?

Based on the mixed-methods analysis aforementioned at the beginning of Sec. 6, we first derive how will instructors interact with our proposed system to deliver synchronous online classes, then we present the findings of how will teachers' behaviors, emotions, attentional and cognitive workloads be influenced by the system.

6.2.1 Usage pattern. We present the findings of how instructors leverage the adaptability of our system and how they interact with the system during the lecture, as well as the underlying reasons for their behaviors.

Finding 1: Most instructors adapt the system to their own needs and preferences. 15 out of 18 instructors reconfigured the interface before class. To name a few, 13 teachers changed the visual form of charts, eight adjusted the reminding mechanism and threshold of the reminder, and four changed the color of charts. In contrast, during the class, fewer (4 out of 18) instructors adjusted the interface on the fly (changing the position of separate views and the size of the in-class window). According to the interviews, teachers prefer to customize the system before class as there is little extra energy and time for them to modify the system during the lecture. From their interaction pattern, we can find different instructors have different habits and preferences in the interface. For example, ten participants chose not to set reminders as they preferred to actively observe students' learning status rather than being prompted to do so. On the contrary, the other eight participants set the reminder because they hoped to be alarmed by the system, worrying that they would be too involved in the lecture and missed important information. Among them, five chose the mode of "always reminding", whereas the other three preferred "reminding once". These observed usage patterns can support the importance of considering instructors' differences when designing online class assistant systems.

Finding 2: Instructors glanced at the sidebar view both intentionally and unconsciously. During the class, viewing (glancing at) the sidebar is an important interaction behavior to analyze. We confirmed most viewing behaviors with participants through co-reviewing the recordings. Overall, all instructors viewed the sidebar during the class, with an average frequency of 26 times (SD=7.1), which is beyond 12 instructors' expectations since they did not realize that they have **Glanced** at the sidebar so many times. This finding is consistent with the "Time spent on student" question (shown in Figure 9 (a)) where instructors spent significantly more time on students with *Glancee* than *EngageClass* (Z=-3.40, p<.001) and *ZoomOnly* (Z=-3.11, p<.01). We conclude 11 kinds of timing and motivation of instructors' behaviors of viewing the sidebar (with the participants covered in each case), shown in Table 3.

Finding 3: Different instructors pay different amounts of attention to different learning statuses. When asking participants what information they focused on during the teaching process, we found that different instructors allocated a different amount of attention to students' different learning statuses (shown in Figure 8). The reason was that everyone had their own consideration criteria. For example, T5 believed that emotion was not important because students' emotion "*may be affected by many factors other*

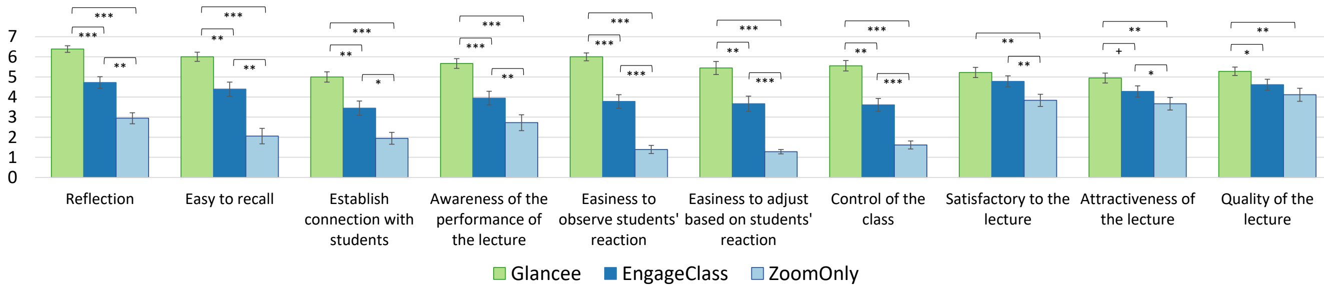


Figure 7: Instructors' ratings on how different feedback systems support their teaching. The error bars indicate standard errors. (+: p<.1; *: p<.05; **: p<.01; ***: p<.001)

Table 3: The timing and reason/motivation of instructors' glancing at the sidebar (mentioned by participants in brackets)

Timing	Reason/Motivation
1. When teaching difficult parts	To ensure students' understanding (T3, T6, T7, T8)
2. After talking for a long time	Worrying students may get bored (T2, T3, T7)
3. After finishing one part or switching slides	To judge students' understanding of the knowledge just taught so as to decide whether to continue or provide a further explanation (T2, T3, T5, T6, T7)
4. When something unexpected happens (e.g., software crashes)	To check whether the students' states get influenced (T1, T2, T5, T7, T14)
5. When being familiar with the content to teach	Having "*energy*" to observe students' learning status when they "*didn't need to focus on the slides*", "*knew what to say next*" or "*played a one-minute video*" (T1, T3, T7, T12)
6. When attracted by visual elements	Being attracted by some obvious visual elements/animations, e.g., "*the reminder appeared*", "*the chart changed obviously*", "*emoji appeared*" (T3, T5, T8, T11, T14, T15, T18)
7. Glancing *unconsciously*	Without specific motivation, just glancing at the sidebar (mentioned by **all** participants)
8. To talk about an important part	To see if the students were "*in good condition*" and "*ready*" to listen (T2, T9, T17)
9. When not feeling confident	Feeling somewhere was not well taught, thus being urgent to look at the sidebar to see "*if students' learning status got worse*" (T4, T7, T14)
10. When being confident in teaching	To seek positive feedback because students should be interested (T10, T12, T13)
11. After making some adjustments	To check whether the adjustment *worked* (mentioned by 16 participants)

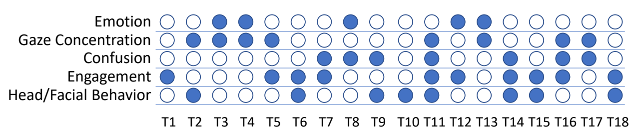


Figure 8: Instructors' attention to different information during the teaching process. The colored circles indicate which views each instructor cared about. We can see clearly that different instructors had different information preferences.

than the lecture itself" while T12 believed that emotion was the most noteworthy information, and he emphasized the importance that "*If students' positive emotions could be activated, students would achieve good learning performance*". This finding is also consistent with our design motivation that **different teachers may have different concerns and preferences**, and thus simply presenting a fixed type of information cannot meet the needs of all teachers.

Furthermore, to our surprise, although we provided the adaptability for instructors to hide any view (learning status), we found that 17 out of 18 teachers kept all views on. However, they mainly focused on about two views, because they *"did not have the energy to perceive too much information at one time"*, which is consistent with our design considerations. Specifically, when asked **why not hide the information they did not pay attention to**, they mainly held three reasons. First, it was the first time they delivered a lecture with *Glancee*, so they wanted to try all views. Second, 15 out of 18 instructors reported that displaying all the information did not cause interference (for details, please see Sec. 6.2.3), so there was no need to turn off any view. Third, 16 out of 18 participants wanted to see as comprehensive students' learning status as possible to "*form a complete perception of the state of students*".

6.2.2 Effects on Instructors' Behavior and Emotions. We analyzed the effects of *Glancee* both on instructors' teaching behaviors and emotions (overall emotions and dynamic changes), and obtained the following findings.

Finding 1: Instructors adjusted their teaching according to student learning status. From the interview, we found that all instructors had made adjustments based on the in-class view. We conclude the common behaviors which happened to at least two participants as well as the causes and examples, shown in Table 4. We also found that some instructors had a multi-round viewing/adjusting pattern. For example, T2 gave an example after he found some students got confused, then he checked the sidebar again and found that there were still some students confused, so he gave a more intuitive example for clarification and moved to the next part until he found no confused students.

Finding 2: Overall, instructors felt more positive about their lecture delivery experiences with *Glancee.* Figure 9 (b) shows instructors' overall emotions when delivering lectures in different conditions. Friedman tests show that different feedback systems have significantly different influences on instructors' feeling engaged ($\chi^2(2)$=11.92, p<.01), frustrated ($\chi^2(2)$=9.64, p<.01), happy ($\chi^2(2)$=10.45, p<.01), hopeful ($\chi^2(2)$=9.67, p<.01), safe ($\chi^2(2)$=9.32, p<.01). Post-hoc tests show that *Glancee* makes instructors significantly more engaged in teaching than *ZoomOnly* (Z=-2.14, p<.05) and has a trend to raise instructors' engagement than *EngageClass* (Z=-1.87, p=.06). And instructors feel significantly less frustrated in *Glancee* than in *ZoomOnly* (Z=-2.83, p<.01). Besides, instructors with *Glancee* feel significantly happier, more hopeful, safer than with *EngageClass* (Z=-1.98, p<.05; Z=-2.40, p<.05; Z=-2.30, p<.05) and *ZoomOnly* (Z=-2.80, p<.01; Z=-2.77, p<.01; Z=-2.95, p<.01). However, no significant difference can be found across three different conditions in terms of instructors' feeling excited, exhausted, overwhelmed, nervous, anxious, or confident.

Finding 3: Instructors' in-situ emotions during the teaching process can be affected by the displayed students' learning status. From the retrospective review and interview with instructors, we found that on the one hand, most (13 out of 18) instructors regarded the students' good learning status as a kind of *"positive feedback"* and a *"recognition of their teaching"*. Therefore, they would feel positive (e.g., confident, delighted, encouraged, sense of achievement, teaching more actively) when seeing students' good learning status, such as students' smiling emoji, positive emotion distribution, head nodding, high engagement, reduced confusion, concentrated gaze, etc. For example, T2, T4, T6 and T11 felt a sense of achievement after making adjustments to their teaching and seeing students' learning status change from poor to good. On the other hand, some instructors' emotions would temporarily become negative (e.g., anxious: T4, 5, 16, 18; depressed: T4, 5, 10, 11, 13, 17; nervous: T2, 13, 18) when seeing students' poor learning status, such as negative emotion distribution, unfocused gaze, continuously low engagement, a sudden drop of engagement, drowsiness, high confusion, head shaking, etc. We found that inexperienced instructors (i.e., participants who are TAs with less teaching experience) were more easily to feel negative. For example, the participants who felt nervous were all TAs. We infer that this may be due to their lack of proficiency and confidence in teaching.

6.2.3 Effects on Instructors' workload, attentional load, and cognitive load. Figure 9 (a) shows instructors' average ratings on the workload, attentional load, and cognitive load questions. Friedman tests show that there are significant differences in instructors' workload ($\chi^2(2)$=11.53, p<.01), attentional load ($\chi^2(2)$=8.40, p<.05) and cognitive load ($\chi^2(2)$=14.00, p<.01) when using different feedback systems to deliver lectures. Post-hoc tests show that compared with *ZoomOnly* (workload: M=2.17, SD=1.25; attentional load: M=2.39, SD=1.61; cognitive load: M=2.17, SD=1.62), *Glancee* (workload: M=3.33, SD=1.61; attentional load: M=3.78, SD=1.66; cognitive load: M=3.56, SD=1.62) induced significantly more workload (Z=-2.12, p<.05), attentional load (Z=-2.32, p<.05) and cognitive load (Z=-2.47, p<.05). However, no significant difference can be found between *Glancee* and *EngageClass* (workload: M=3.28, SD=1.49; attentional load: M=3.28, SD=1.36; cognitive load: M=2.94, SD=1.47), which reveals that the designed multi-information interface did not introduce more burden on users compared with single-information interface. Overall, it is intuitive that *ZoomOnly* introduced the least workload, attentional load and cognitive load, because without students turning on the camera, instructors "*could not observe any students' learning status except for the slides*" and thus "*did not need to understand any information*". We also obtained some interesting findings from participants' qualitative feedback.

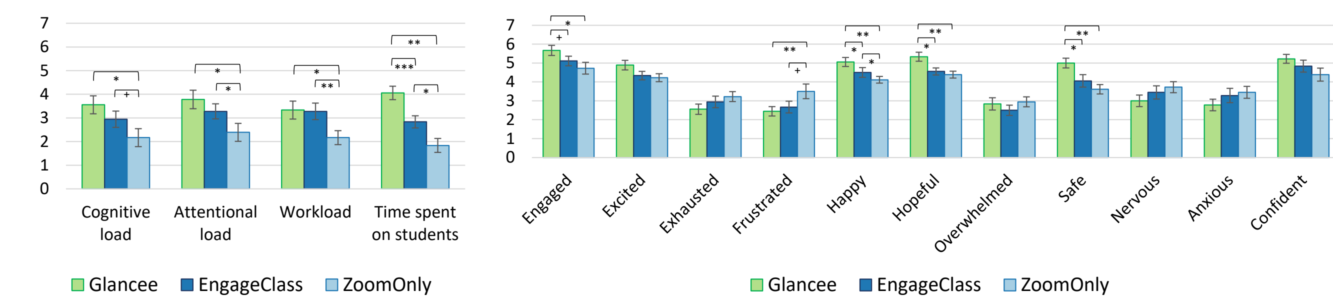
Finding 1. The workload, attentional load and cognitive load are negligible for instructors. Almost all instructors thought the burden brought by the system was very light. The reasons can be categorized as follows: 1) It was the instructors' initiative to observe the state of students instead of a "*have to*" behavior, so they "*didn't think it was a burden*"(T3, 8, 9, 10, 12, 14, 17). 2) Instructors only needed to **glance at** the sidebar without twisting their heads between two screens, so they didn't spend much time and effort. Participants thought it was "*natural to switch between slides and the sidebar*" (T5, 9, 17), and their gaze could "*immediately return to the slide.*" (T4). 3) Instructors usually looked at the sidebar when they didn't have to focus on slides. At this time, their attention and cognition bandwidth was "*relatively free*" to process information (T6, 15). 4) Usually, teachers only chose to see one or two views that they thought important and the views were "*visually simple enough*", which did not cause much attentional and cognitive burden. This finding is consistent with users' ratings on their workload (M=3.33), attentional load (M=3.78), and cognitive load (M=3.56), which are all lower than the middle bar (i.e., 4) in the 7-point Likert questions.

Finding 2. A little load was perceived to bring more advantages than disadvantages, just like observing the students in the offline classroom. All instructor participants were willing to spend time observing students' learning status and 17 participants expressed that compared with the extra load, they cared more about whether they could establish a connection with students. For example, T16 said "*The benefit this tool brought to me was far greater than the so-called burden. It's like an assistant helping me convey the state of the class.*" and T7 said if he was not provided with the sidebar, he would be more likely to "*feel anxious*" because it would be "*a situation beyond his control*". Many teachers compared observing the sidebar to observing students in the classroom and found them similar. T11 mentioned "*A moderate load was what exactly a teacher wants. Even in a physical classroom, I would spend time reading the classroom*". And some instructors reported that they "*just turned the process of looking at students offline into looking at the sidebar online*" (T3) and they "*actually regarded this interface [the sidebar view] as a window to see students*"(T14).

Finding 3. Cognitive load was more likely to occur when the displayed student status was inconsistent with the teacher's expectation. At this time, the teacher needed to think about what happened and reflect on how to adjust the teaching. For example, T4 said "*When students had negative emotions or their attention was not very focused, my cognitive load would increase, because I would reflect on whether I didn't explain the content clearly.*" T15 said "*When students didn't pay much attention, I would think about how to deal with it, which brought me a low cognitive load.*"

Table 4: Instructors' behaviors during the teaching process with the proposed *Glancee*.

Behavior	Triggered by	Example
1. Giving examples	High confusion (T1, T2, T7, T8, T12), head shaking (T14), frowning (T7, T12)	T2 introduced an example in students' daily life to help them understand fabrication.
2. Showing demos	Low engagement (T3, T5, T15), high confusion (T9, T14, T17), negative emotion (T10)	T3 opened a demo program to illustrate how gestures were recognized.
3. Moving to the attractive part	Low engagement (T5, T6), negative emotion (T3, T12), drowsiness (T9)	T5 moved to the next page with vivid figures to attract students' attention.
4. Moving to key points	Low engagement (T7, T15), high confusion (T9, T16)	T7 skipped technical details which are "*not the focus of the course*".
5. Injecting humor or a relaxing topic	Negative emotion (T12, T13), low engagement (T7, T12), distracted gaze (T3, T17)	T12 told a joke, and T17 talked about some hot news to activate the class.
6. Highlighting or drawing on the slide	High confusion (T3, T14), distracted gaze (T17)	T14 and T17 used a pen [in Zoom] to demonstrate the key steps in a formula.
7. Speeding up the speech	Low engagement (T3, T5), negative emotion (T4, T12)	T5 accelerated his speech when the content could not be skipped but students' status was poor.
8. Slowing down or injecting pauses	High confusion (T6, T7, T8, T9, T11, T16, T17), head shaking emoji (T14)	T6 slowed down her speech to "*give students time to digest the content.*".
9. Repeating and providing further clarification	Low engagement (T1, T11), high confusion (T6, T14, T15)	T14 repeated the content just taught when the content was important.
10. Asking for students' confusion points	High confusion (T6, T8, T9, T11, T14, T18), low engagement (T17)	T18 directly asked "*Any questions?*", T9 asked questions to check whether the students had understood.
11. Reminding students to pay attention	Low engagement (T11, T17), speaking emoji (T11), drowsiness (T17)	T11 encouraged students to pay attention.
12. Reviewing whether something taught is wrong	High confusion (T3, T6), head shaking emoji (T14)	T14 spent time to check the content after observing a head-shaking emoji.



(a) (b)

Figure 9: The effects of different feedback systems on instructors. (a) The effects on instructors' cognitive load, attentional load, workload, and time spent on observing students. (b) The effects on instructors' emotions and related feelings. The error bars indicate standard errors. (+: *p*<.1; *: *p*<.05; **: *p*<.01; *: *p*<.001)**

Finding 4. Instructors' cognitive load decreased as they became more familiar with the system. There was a learning curve for some teachers. In the interview, they reported that they "*felt some cognitive load at the beginning, but the cognitive load was decreased after perceiving the charts several times*". For example, in the beginning, they were "*not familiar with mapping each chart to the meaning behind it*" (T17), but as they looked at the views several times, it became "*easy to establish a mapping*" (T16, T18). Another reason is that instructors did not fully master the functions of the interface in the tutorial process, so they need to be further familiar with it in real usage (T13, T18).

6.3 RQ5: How will instructors trust and collaborate with such an AI-empowered system?

6.3.1 How will instructors trust Glancee *and what might affect their trust?* Figure 10 (a) shows participants' average ratings in the 7-point Likert trust-related questions of the post-task questionnaire. The average perceived accuracy of the system is 5.61 (*SD*=0.18), and the overall trust in the system is 5.67 (*SD*=0.18), which shows that participants tend to feel that the students' learning status conveyed by the system is accurate and believe in the system, although not completely. Moreover, for each type of students' learning status, teachers showed different levels of trust. These results were further supported by the qualitative findings from the interview.

Finding 1. Instructors tended to believe in the system but not fully. All participants rated their trust over the middle bar (i.e., 4) of the 7-point Likert question. They believed in the system because they "*couldn't find a strong reason not to trust this system*" (T14) and "*no obviously questionable information was found*" (T2, T11, T12, T17). This is consistent with observations in psychology literature on *Truth-Default Theory* [50]. However, due to the lack of ground truth (the students' videos), they could not fully trust the system (T6, T16, T18). Furthermore, some factors also prevented them from fully trusting the system (detailed below).

Finding 2. When some instructors encountered situations that did not meet their expectations, they would first doubt themselves rather than the system. Some instructors will not doubt the system unless the displayed information is abnormal. For example, T7 said "*When I found the content very attractive, but the students showed a poor engagement, I did not doubt the system first. Instead, I doubted myself that maybe I didn't express some details clearly.*" T13 said "*When I told jokes, students showed negative emotions. I wondered if I said something inappropriate.*"

Finding 3. Instructors' trust in different learning status measures is independent of each other. Instructors' distrust of some learning statuses will not affect their trust in other learning statuses. For example, T2 said "*Even if only one view in the interface could be trusted, I thought I could continue to use this system. I would automatically ignore the information I didn't believe.*" T13 said "*I only trusted in what [information] I thought important.*" This finding is consistent with what is shown in Figure 10 (b) where instructors have varying trust levels on different information.

From the interview, we find that there are several factors affecting instructors' trust.

Factor 1. Whether the displayed students' learning status is as expected. 12 participants mentioned that *if the displayed students' learning status meets their expectations, they will strongly trust the information.* For example, some instructors found the low engagement reliable because "*I had talked too long, and the audiences might be tired*" (T2), "*I was talking about the boring code part*" (T3). Similarly, T7 trusted gaze and emotions when he found "*the value of gaze suddenly fell down and students' emotion became negative when my Powerpoint crashed*". T8 particularly believed in confusion because he thought "*the content was difficult on page 8 and students really showed confusion*" and the students became not confused after he "*gave some examples*". On the contrary, *if the displayed students' learning status mismatches instructors' expectations, they will doubt the system.* For example, T3 doubted the confusion as he found students confused when he was "*teaching a not difficult part.*"

Factor 2. Frequency of chart changes. On one hand, if the charts changed too frequently, instructors might doubt it. For example, T3 said "*Students' confusion changed too frequently (kept shifting), which made me doubt it.*" T8 said "*I found it strange that engagement changed dramatically in the same slide.*" On the other hand, if the charts kept still for a long time, it would also make instructors doubt it. For example, T10 said "*I saw that the students' state had not changed too much in the past five minutes. I wondered whether it was a problem of this system.*"

Factor 3. Whether the displayed students' learning status is positive or negative. First, instructors tend to doubt the negative information even it meets their expectations. For example, T1 mentioned "*There were many negative emotions among students which should be caused by something wrong with my video sound, but I was still a little skeptical.*" Second, instructors tend to believe the positive information even it does not meet their expectations. For example, T5 said "*I believed in gaze, because it was always 100% concentrated. Although strangely high, it should be true.*"

Factor 4. Instructors' intrinsic perceptions of the sensing techniques. Their intrinsic perception played a big role in their trust. For example, T1 said "*I didn't believe in emotion very much because I had learned about emotion detection algorithms before. In my impression, these algorithms were not perfect enough.*" T2 stated "*I believed in gaze, because gaze was an intuitive measurement, and the algorithm might not be error-prone.*" And T17 mentioned "*Because in the news, AI-based affective recognition was pretty mature and popular, so I believed in the system.*" (Note that T17 does not major in computer science, and he just guessed and judged the potential algorithm based on his experience.)

Factor 5. Instructors' perceived importance of the learning status. Teachers prefer to believe what they think is important. For example, T6 said "*I didn't trust gaze, because staring at the screen didn't necessarily mean that students were listening carefully, so I subconsciously thought the displayed gaze was inaccurate.*" T8 said "*Except for the confusion which I paid more attention to, other information might not be so accurate.*"

6.3.2 Teacher-AI Collaboration.

Besides the trust issue, we got some interesting findings of how instructors collaborated with the AI-empowered assistant system from their open-ended feedback during the interview.

Finding 1. Teachers would take absolute control during the interaction. From the qualitative feedback, we found that all teachers had absolute control of their behaviors and the teaching process such as 1) whether to look at the sidebar or not, 2) whether to trust the information and which one to trust, 3) whether to make adjustments according to the displayed student learning status and how to adjust. Teachers utilized the system in combination with their own understanding of the class instead of merely relying on the AI-predicted information.

Finding 2. The system is like an evaluator, encouraging instructors to improve their teaching. A lot of instructors expressed that they wanted to achieve a better "score" in the tool. For example, T3 said "*There was a quantitative metric telling you how your teaching performance was. I couldn't help trying to get high marks.*" T11 stated "*If I can use this system in the future, I will constantly adjust my teaching in later classes until the line [historical learning status] in the post-class view reaches a good level.*"

Finding 3. Sometimes the system can serve as a mentor for instructors, helping identify problems that they have not realized before. For example, T3 mentioned "*I always thought I taught the principle of gesture recognition clearly. Today, I just found that students didn't understand it, so I would improve my slides after the class.*" And T16 thanked the system "*I have taught this course for many years, but my slides changed little since the first semester due to*

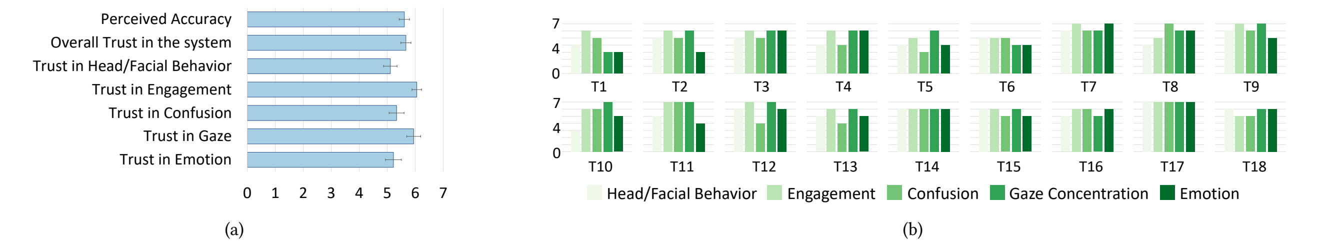


Figure 10: Instructors' trust in *Glancee*. (a) Instructors' perceived accuracy of the system, their overall trust in the system, and the average trust in each kind of information. (b) Instructors' trust levels (in 7-point Likert) on different information.

the lack of intuitive feedback. [With this system], I could continuously reflect, adjust my teaching methods and contents."

7 DISCUSSION

This work looks into the problem of instructors' difficulty in observing students' learning status in synchronous online classes. Overall, we revealed the individual differences in learning status information seeking among instructors, explored the design of *Glancee*, and investigated its effects on instructors' teaching behaviors, emotions, attentional and cognitive workloads, as well as trust. In this section, we discuss the overarching design issues of *Glancee* and build upon our key findings to propose design recommendations for general online teaching assistant systems.

7.1 Design Issues and Design Recommendations

7.1.1 ***Negative emotions caused by the transparency of students' learning status.*** Overall, our system evoked statistically more positive emotions (e.g., more engaged, happier, and less frustrated) in instructors than the two baselines (Finding 2 in Sec. 6.2.2). However, during the teaching process, some instructors also experienced anxiety, depression, or tension due to students' poor learning statuses prompted by *Glancee* (Finding 3 in Sec. 6.2.2). From the interview, we found that such negative emotions just existed temporarily because instructors would adjust their teaching to improve students' learning statuses. Nevertheless, we acknowledge that the negative emotion is not necessarily a minor issue, especially for inexperienced instructors who are more likely to be nervous about teaching, because the negative information may affect their teaching confidence and performance [63]. In our experiment, the instructors who reported getting stressed over the negative feedback were all TAs with less teaching experience.

Design Recommendation 1: Avoiding bringing negative emotions to instructors. Designers can consider adopting a non-judgmental information display [51, 80] to deliver gentle information to teachers when students are in a poor state. What's more, designers can utilize the proposed *Customization of Reminder* function (described in Sec. 4.3.2) to suggest a lower alert threshold for less experienced instructors so that the interface will be more "tolerant".

7.1.2 ***Instructors' (over)trust in the imperfect AI algorithms.*** Our results show that teachers had complete control over how they delivered the lectures based on their own teaching abilities and intuition, rather than completely relying on the students' learning statuses presented by the interface (Finding 1 in Sec. 6.3.2). Besides, when the displayed learner states did not match their expectations, they tended to doubt the system (Factor 1 in Sec. 6.3.1). However, we found that some instructors had a tendency to over-trust this system if they found that the system information generally aligned with what they anticipated. Once in this mindset, when seeing poor student states, these instructors would first doubt their own teaching instead of questioning system accuracy (Finding 2 in Sec. 6.3.1). Since computer vision-based algorithms are far from perfect [54], over-trusting the system could undermine instructors' teaching confidence, especially when a good learning state is mistakenly recognized as a poor one. This may cause unnecessary mental stress in teachers and disrupt their original teaching flow.

Design Recommendation 2: Calibrating instructors' trust in AI-assisted systems. On the one hand, designers can enhance the transparency of algorithms, such as visualizing the confidence/uncertainty [10] or providing explanations [104]. On the other hand, we suggest designers guide instructors to establish an accurate mental model of the system [7, 34]. For example, designers can explain the main principle of the utilized algorithm in a way that teachers, even non-computing professionals, can understand [18, 73], and designers can inform instructors of the capabilities and any possible pitfalls of the system [72, 89]. It will facilitate maintaining a correct perception and appropriate trust in the system so that teachers can leverage the system's support ideally without over-trust or under-trust [49, 56, 104].

7.1.3 ***Design choices of implicit and explicit feedback.*** We used computer vision-based methods to implicitly detect the learning status of students, which brought in many benefits such as continuous real-time feedback and negligible distraction for instructors and students [68, 85]. This, as shown in our study, allowed instructors to establish connections with students and make real-time adjustments (Finding 1 in Sec. 6.2.2). However, we acknowledge that implicit sensing methods require careful considerations of privacy issues. Besides, we should note that the implicit method is not perfect and can sometimes miss or mistakenly detect some important states [86, 92]. More explicit feedback methods (e.g., self reports) may offer more accurate feedback and may be preferable for those who are more concerned about privacy [54]. Nevertheless, explicit feedback would increase cognitive workload and distraction for the audience [68, 85], and audiences might forget to provide feedback when they are too attentive or distracted [54].

Design Recommendation 3: Deciding feedback method based on specific application scenarios. For example, an implicit method in a formal lecture class may be more appropriate, while in an interactive or discussion class, an explicit method could be more suitable. In addition, we suggest designers explore an integrated interface to utilize both implicit and explicit methods to cover various online teaching scenarios and different users and allow users to decide which type of feedback to utilize based on specific needs and scenarios.

7.1.4 *Instructors' differences and preferences in online teaching*.

Instructors have their own teaching goals, styles and habits [57]. Similarly, our survey and experiment reveal individual differences among the instructors in terms of their information preferences (Sec. 3.2.1), usage patterns of adaptable features (Finding 1 in Sec. 6.2.1), attention allocation strategy (Finding 3 in Sec. 6.2.1), and level of trust on the AI system (Finding 3 in Sec. 6.3.1), etc. It is thus hard, if not impossible, to design a "one-size-fits-all" teaching-support system that can meet the needs of every teacher.

Design Recommendation 4: Empowering instructors with adaptability and control over the system. We suggest designers provide instructors with the adaptability of the system, such as the flexibility to customize the interface display, whether and how to use a specific feature, etc. On the other hand, we suggest designers respect instructors' control over their teaching process. The system should play an assistant role. Even if the system exhibits behavior that does not meet instructors' preferences, they can ignore it and follow their own teaching pace.

7.1.5 *Instructors' attentional and cognitive workloads during the teaching process*.

To facilitate teachers' real-time customization of the interface, we provide some interactive supports for the in-class view. However, we observed that 14 out of 18 instructors did not adjust the interface at all during the class (Finding 1 in Sec. 6.2.1) because customizing the system while teaching requires complete attention transfer from content delivery to the interface. It will be more feasible for them to only perform *light* interactions, such as glancing at the sidebar in our case. From the user study, we find that instructors appreciated the visually simple but effective sidebar design, because they can easily distill certain information from the sidebar and only focus on what they are concerned about, which leads to moderate cognitive load. Besides, they can easily observe and understand the state of students through "unconscious" glances without switching their gaze frequently between multiple screens (Finding 1 in Sec. 6.2.3).

Design Recommendation 5: Minimizing the burden on instructors. When designing teaching support tools, especially for real-time classes, designers are recommended to take measures to minimize the burden on instructors, such as utilizing intuitive information display [85] to reduce cognitive load, minimizing instructors' interaction workload (e.g., button-clicking, head-turning), so that the supportive features will not disrupt the original teaching flow.

Design Recommendation 6: Avoiding excessive interaction *during* real-time class. For real-time teaching, we suggest designers arrange time-consuming interaction tasks before/after class or during the break time, instead of on the fly, e.g., prompt instructors to make necessary configurations on the system when playing a video.

7.1.6 *Students' privacy concerns of providing learning status information*.

From the exploratory survey (Sec. 3.2.2) and interviews with students (Appendix B), we found that the student participants really need and appreciate the feature of anonymity and the privacy concern greatly affects their willingness to deploy the system. We suggest that future design should take students' privacy concerns into consideration when collecting students' personal data.

Design Recommendation 7: Providing students with control of the system. During the experiment, some students suggested making *Glancee* as a toggle button in Zoom, such as "video" button, so that they can choose whether to turn it on freely. Thus, when designing similar learning status feedback systems, the interface can empower students with the control of whether to contribute and which types of learning status to contribute to instructors [54].

Design Recommendation 8: Utilizing the anonymity mechanism. Previous studies [17] have found that the anonymity of Danmaku encouraged more students to participate in class. And students surveyed in our work also revealed the importance of anonymous feedback. We suggest that designers provide anonymous feedback channels for students to enhance their initiative and seek for a combination of real-name systems and anonymity.

7.2 Limitations and Future Work

First, the performance of the computer vision-based algorithm could be negatively influenced by video backgrounds, lighting conditions, and camera angles [54]. On the one hand, we can adopt more robust sensing algorithms. On the other hand, we can communicate the uncertainty of the detection results to instructors. To achieve the latter, we need to consider the cognitive load and limited attention resources of teachers when introducing additional information.

Second, in this paper, we only test the system in a formal lecture class where instructors introduce knowledge based on slides. However, there are some other kinds of classes, such as discussion classes, interactive classes, flipped classes [12, 35], where instructors' teaching goals and roles could be different from a formal lecture class. In the next step, we will explore the context and challenges in other types of online classes, and adapt our system to the new user needs.

Third, from interviews, we found that some instructors treated this system as an evaluation metric, and they wanted to improve their teaching methods to improve their scores. So, it can be very interesting to deploy a long-term experiment to explore how the system will affect the behavior, emotions, and cognitive burden of teachers as time goes on, as well as whether the quality of instructors' teaching will be improved during the long-term use of the system.

In addition, we foresee that besides purely online/offline mode, the mixed-mode (hybrid learning) is likely to become a trend in the future [91], where some students are in the physical classroom and some are remotely attending. In this situation, instructors' attention needs to switch between the offline and online audience, which will lead to a great burden on instructors and damage students' learning experiences. In the future, we could investigate how to embed our interface in the physical teaching environment via immersive visualization [27] or AR techniques [39, 99].

8 CONCLUSION

We propose *Glancee*, an adaptable system to address instructors' difficulty to observe students' learning status due to students' unwillingness to show their videos. Specifically, we mitigate the gap that lack of empirical investigation on instructors' preferences and lack of exploration of designing adaptable systems to meet the needs of individual instructors. We summarize our contributions as follows. First, we conducted two exploratory surveys to understand instructors' actual needs and students' concerns of conveying students' learning status to instructors in synchronous online classes. Second, we designed and implemented *Glancee*, which provides instructors with a real-time students' learning status display with an adaptable interface. Third, we conducted a within-subject user study with 18 instructors to verify the effectiveness of our design, explore how instructors will interact with the system in actual synchronous online classes, and comprehensively investigate the effects of the system on instructors. Our results highlight instructors' differences in information seeking and usage patterns, which reveals the importance of the adaptable interface design. Besides, we conclude findings of the influence on instructors' behaviors, emotions, attentional and cognitive workloads, as well as trust and collaboration issues in practical online teaching with *Glancee*. We believe our exploratory survey (RQ1-2) and user study (RQ3-5) illuminate one possible direction for designing instructor-adaptable systems to enhance instructor-student connections in synchronous online classes.

ACKNOWLEDGMENTS

The authors would like to thank the reviewers for their insightful feedback and thank the participants of our studies for their time. This work is supported by the Hong Kong General Research Fund (GRF) with grant No. 16204819.

REFERENCES

[1] Karan Ahuja, Dohyun Kim, Franceska Xhakaj, Virag Varga, Anne Xie, Stanley Zhang, Jay Eric Townsend, Chris Harrison, Amy Ogan, and Yuvraj Agarwal. 2019. EduSense: Practical classroom sensing at Scale. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 3, 3 (2019), 1–26.
[2] Karan Ahuja, Deval Shah, Sujeath Pareddy, Franceska Xhakaj, Amy Ogan, Yuvraj Agarwal, and Chris Harrison. 2021. Classroom Digital Twins with Instrumentation-Free Gaze Tracking. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. 1–9.
[3] Misna Ali, Shahsad Abdullah, CS Raizal, KF Rohith, and Varun G Menon. 2019. A novel and efficient real time driver Fatigue and Yawn detection-alert system. In *2019 3rd International Conference on Trends in Electronics and Informatics (ICOEI)*. IEEE, 687–691.
[4] Pengcheng An, Saskia Bakker, Sara Ordanovski, Ruurd Taconis, Chris LE Paffen, and Berry Eggen. 2019. Unobtrusively enhancing reflection-in-action of teachers through spatially distributed ambient information. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. 1–14.
[5] Microsoft Face API. 2021. An AI service that analyzes faces in images. https://azure.microsoft.com/en-us/services/cognitive-services/face/. Accessed September 9, 2021.
[6] Tadas Baltrusaitis, Amir Zadeh, Yao Chong Lim, and Louis-Philippe Morency. 2018. Openface 2.0: Facial behavior analysis toolkit. In *2018 13th IEEE international conference on automatic face & gesture recognition (FG 2018)*. IEEE, 59–66.
[7] Gagan Bansal, Tongshuang Wu, Joyce Zhou, Raymond Fok, Besmira Nushi, Ece Kamar, Marco Tulio Ribeiro, and Daniel Weld. 2021. Does the whole exceed its parts? the effect of ai explanations on complementary team performance. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. 1–16.
[8] Emad Barsoum, Cha Zhang, Cristian Canton Ferrer, and Zhengyou Zhang. 2016. Training deep networks for facial expression recognition with crowd-sourced label distribution. In *Proceedings of the 18th ACM International Conference on Multimodal Interaction*. 279–283.
[9] Roman Bednarik, Shahram Eivazi, and Michal Hradis. 2012. Gaze and conversational engagement in multiparty video conversation: an annotation scheme and classification of high and low levels of engagement. In *Proceedings of the 4th workshop on eye gaze in intelligent human machine interaction*. 1–6.
[10] Umang Bhatt, Javier Antorán, Yunfeng Zhang, Q Vera Liao, Prasanna Sattigeri, Riccardo Fogliato, Gabrielle Melançon, Ranganath Krishnan, Jason Stanley, Omesh Tickoo, et al. 2021. Uncertainty as a form of transparency: Measuring, communicating, and using uncertainty. In *Proceedings of the 2021 AAAI/ACM Conference on AI, Ethics, and Society*. 401–413.
[11] Timothy Bickmore, Everlyne Kimani, Ameneh Shamekhi, Prasanth Murali, Dhaval Parmar, and Ha Trinh. 2021. Virtual agents as supporting media for scientific presentations. *Journal on Multimodal User Interfaces* 15, 2 (2021), 131–146.
[12] Charles C Bonwell and James A Eison. 1991. *Active Learning: Creating Excitement in the Classroom. 1991 ASHE-ERIC Higher Education Reports.* ERIC.
[13] Virginia Braun and Victoria Clarke. 2006. Using thematic analysis in psychology. *Qualitative research in psychology* 3, 2 (2006), 77–101.
[14] Zhe Cao, Gines Hidalgo, Tomas Simon, Shih-En Wei, and Yaser Sheikh. 2019. OpenPose: realtime multi-person 2D pose estimation using Part Affinity Fields. *IEEE transactions on pattern analysis and machine intelligence* 43, 1 (2019), 172–186.
[15] AT Chamillard. 2011. Using a student response system in CS1 and CS2. In *Proceedings of the 42nd ACM technical symposium on Computer science education*. 299–304.
[16] Xinyue Chen, Si Chen, Xu Wang, and Yun Huang. 2021. " I was afraid, but now I enjoy being a streamer!" Understanding the Challenges and Prospects of Using Live Streaming for Online Education. *Proceedings of the ACM on Human-Computer Interaction* 4, CSCW3 (2021), 1–32.
[17] Zhilong Chen, Hancheng Cao, Yuting Deng, Xuan Gao, Jinghua Piao, Fengli Xu, Yu Zhang, and Yong Li. 2021. Learning from Home: A Mixed-Methods Analysis of Live Streaming Based Remote Education Experience in Chinese Colleges during the COVID-19 Pandemic. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. 1–16.
[18] Hao-Fei Cheng, Ruotong Wang, Zheng Zhang, Fiona O'Connell, Terrance Gray, F Maxwell Harper, and Haiyi Zhu. 2019. Explaining decision-making algorithms through UI: Strategies to help non-expert stakeholders. In *Proceedings of the 2019 chi conference on human factors in computing systems*. 1–12.
[19] Mathieu Chollet, Torsten Wörtwein, Louis-Philippe Morency, Ari Shapiro, and Stefan Scherer. 2015. Exploring feedback strategies to improve public speaking: an interactive virtual audience framework. In *Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing*. 1143–1154.
[20] Scotty Craig, Arthur Graesser, Jeremiah Sullins, and Barry Gholson. 2004. Affect and learning: an exploratory look into the role of affect in learning with AutoTutor. *Journal of educational media* 29, 3 (2004), 241–250.
[21] Didan Deng, Zhaokang Chen, and Bertram E Shi. [n.d.]. Multitask Emotion Recognition with Incomplete Labels. In *15th IEEE International Conference on Automatic Face and Gesture Recognition (FG 2020)(FG)*. IEEE Computer Society, 828–835.
[22] Mandalapu Sarada Devi and Preeti R Bajaj. 2008. Driver fatigue detection based on eye tracking. In *2008 First International Conference on Emerging Trends in Engineering and Technology*. IEEE, 649–652.
[23] M Ali Akber Dewan, Mahbub Murshed, and Fuhua Lin. 2019. Engagement detection in online learning: a review. *Smart Learning Environments* 6, 1 (2019), 1–20.
[24] Abhinav Dhall, Garima Sharma, Roland Goecke, and Tom Gedeon. 2020. Emotiw 2020: Driver gaze, group emotion, student engagement and physiological signal based challenges. In *Proceedings of the 2020 International Conference on Multimodal Interaction*. 784–789.
[25] Elena Di Lascio, Shkurta Gashi, and Silvia Santini. 2018. Unobtrusive assessment of students' emotional engagement during lectures using electrodermal activity sensors. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 2, 3 (2018), 1–21.
[26] Sidney D'Mello, Andrew Olney, Claire Williams, and Patrick Hays. 2012. Gaze tutor: A gaze-reactive intelligent tutoring system. *International Journal of human-computer studies* 70, 5 (2012), 377–398.
[27] Ciro Donalek, S George Djorgovski, Alex Cioc, Anwell Wang, Jerry Zhang, Elizabeth Lawler, Stacy Yeh, Ashish Mahabal, Matthew Graham, Andrew Drake, et al. 2014. Immersive and collaborative data visualization using virtual reality

platforms. In *2014 IEEE International Conference on Big Data (Big Data)*. IEEE, 609–614.
[28] Panorama Education. 2019. User Guide: Panorama Teacher and Staff Survey.
[29] Paul Ekman and Erika L Rosenberg. 1997. *What the face reveals: Basic and applied studies of spontaneous expression using the Facial Action Coding System (FACS)*. Oxford University Press, USA.
[30] Emily K Faulconer, J Griffith, Beverly Wood, S Acharyya, and D Roberts. 2018. A comparison of online, video synchronous, and traditional learning modes for an introductory undergraduate physics course. *Journal of Science Education and Technology* 27, 5 (2018), 404–411.
[31] Anne C Frenzel, Reinhard Pekrun, Thomas Goetz, Lia M Daniels, Tracy L Durksen, Betty Becker-Kurz, and Robert M Klassen. 2016. Measuring teachers' enjoyment, anger, and anxiety: The Teacher Emotions Scales (TES). *Contemporary Educational Psychology* 46 (2016), 148–163.
[32] Katsuya Fujii, Plivelic Marian, Dav Clark, Yoshi Okamoto, and Jun Rekimoto. 2018. Sync class: Visualization system for in-class student synchronization. In *Proceedings of the 9th Augmented Human International Conference*. 1–8.
[33] Nan Gao, Wei Shao, Mohammad Saiedur Rahaman, and Flora D Salim. 2020. n-Gage: Predicting in-class Emotional, Behavioural and Cognitive Engagement in the Wild. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 4, 3 (2020), 1–26.
[34] Katy Ilonka Gero, Zahra Ashktorab, Casey Dugan, Qian Pan, James Johnson, Werner Geyer, Maria Ruiz, Sarah Miller, David R Millen, Murray Campbell, et al. 2020. Mental Models of AI Agents in a Cooperative Game Setting. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. 1–12.
[35] Elena L Glassman, Juho Kim, Andrés Monroy-Hernández, and Meredith Ringel Morris. 2015. Mudslide: A spatially anchored census of student confusion for online lecture videos. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*. 1555–1564.
[36] Sandra G Hart and Lowell E Staveland. 1988. Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. In *Advances in psychology*. Vol. 52. Elsevier, 139–183.
[37] Mariam Hassib, Stefan Schneegass, Philipp Eiglsperger, Niels Henze, Albrecht Schmidt, and Florian Alt. 2017. EngageMeter: A system for implicit audience engagement sensing using electroencephalography. In *Proceedings of the 2017 Chi conference on human factors in computing systems*. 5114–5119.
[38] Danah Henriksen, Edwin Creely, and Michael Henderson. 2020. Folk pedagogies for teacher transitions: Approaches to synchronous online learning in the wake of COVID-19. *Journal of Technology and Teacher Education* 28, 2 (2020), 201–209.
[39] Kenneth Holstein, Bruce M McLaren, and Vincent Aleven. 2018. Student learning benefits of a mixed-reality teacher awareness tool in AI-enhanced classrooms. In *International conference on artificial intelligence in education*. Springer, 154–168.
[40] Hsiu-Fang Hsieh and Sarah E Shannon. 2005. Three approaches to qualitative content analysis. *Qualitative health research* 15, 9 (2005), 1277–1288.
[41] Stephen Hutt, Kristina Krasich, James R. Brockmole, and Sidney K. D'Mello. 2021. Breaking out of the Lab: Mitigating Mind Wandering with Gaze-Based Attention-Aware Technology in Classrooms. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. 1–14.
[42] Stephen Hutt, Caitlin Mills, Nigel Bosch, Kristina Krasich, James Brockmole, and Sidney D'mello. 2017. " Out of the Fr-Eye-ing Pan" Towards Gaze-Based Models of Attention during Learning with Technology in the Classroom. In *Proceedings of the 25th Conference on User Modeling, Adaptation and Personalization*. 94–103.
[43] Amanjot Kaur, Aamir Mustafa, Love Mehta, and Abhinav Dhall. 2018. Prediction and localization of student engagement in the wild. In *2018 Digital Image Computing: Techniques and Applications (DICTA)*. IEEE, 1–8.
[44] Davis E King. 2009. Dlib-ml: A machine learning toolkit. *The Journal of Machine Learning Research* 10 (2009), 1755–1758.
[45] Martin Koestinger, Paul Wohlhart, Peter M Roth, and Horst Bischof. 2011. Annotated facial landmarks in the wild: A large-scale, real-world database for facial landmark localization. In *2011 IEEE international conference on computer vision workshops (ICCV workshops)*. IEEE, 2144–2151.
[46] Nataliya Kosmyna and Pattie Maes. 2019. AttentivU: an EEG-based closed-loop biofeedback system for real-time monitoring and improvement of engagement for personalized learning. *Sensors* 19, 23 (2019), 5200.
[47] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. 2012. Imagenet classification with deep convolutional neural networks. *Advances in neural information processing systems* 25 (2012), 1097–1105.
[48] Denise G Kutnick and David A Joyner. 2019. Synchronous at scale: investigation and implementation of a semi-synchronous online lecture platform. In *Proceedings of the Sixth (2019) ACM Conference on Learning@ Scale*. 1–4.
[49] John D Lee. 2008. Review of a pivotal Human Factors article:"Humans and automation: use, misuse, disuse, abuse". *Human Factors* 50, 3 (2008), 404–410.
[50] Timothy R Levine. 2014. Truth-default theory (TDT) a theory of human deception and deception detection. *Journal of Language and Social Psychology* 33, 4 (2014), 378–392.
[51] James J Lin, Lena Mamykina, Silvia Lindtner, Gregory Delajoux, and Henry B Strub. 2006. Fish'n'Steps: Encouraging physical activity with an interactive computer game. In *International conference on ubiquitous computing*. Springer, 261–278.
[52] Sampada Marathe and S Shyam Sundar. 2011. What drives customization? Control or identity?. In *Proceedings of the SIGCHI conference on human factors in computing systems*. 781–790.
[53] Daniel McDuff, Kael Rowan, Piali Choudhury, Jessica Wolk, ThuVan Pham, and Mary Czerwinski. 2019. A multimodal emotion sensing platform for building emotion-aware applications. *arXiv preprint arXiv:1903.12133* (2019).
[54] Prasanth Murali, Javier Hernandez, Daniel McDuff, Kael Rowan, Jina Suh, and Mary Czerwinski. 2021. AffectiveSpotlight: Facilitating the Communication of Affective Responses from Audience Members during Online Presentations. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. 1–13.
[55] Prasanth Murali, Lazlo Ring, Ha Trinh, Reza Asadi, and Timothy Bickmore. 2018. Speaker hand-offs in collaborative human-agent oral presentations. In *Proceedings of the 18th International Conference on Intelligent Virtual Agents*. 153–158.
[56] Kazuo Okamura and Seiji Yamada. 2020. Adaptive trust calibration for human-AI collaboration. *Plos one* 15, 2 (2020), e0229132.
[57] Marie-Christine Opdenakker and Jan Van Damme. 2006. Teacher characteristics and teaching styles as effectiveness enhancing factors of classroom practice. *Teaching and teacher education* 22, 1 (2006), 1–21.
[58] OpenCV. 2021. OpenCV. https://opencv.org/. Accessed September 9, 2021.
[59] Dhaval Parmar and Timothy Bickmore. 2020. Making It Personal: Addressing Individual Audience Members in Oral Presentations Using Augmented Reality. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 4, 2 (2020), 1–22.
[60] Reinhard Pekrun. 2014. Emotions and learning. *Educational practices series* 24, 1 (2014), 1–31.
[61] 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.
[62] Amber Phelps and Dimitrios Vlachopoulos. 2020. Successful transition to synchronous learning environments in distance education: A research on entry-level synchronous facilitator competencies. *Education and Information Technologies* 25, 3 (2020), 1511–1527.
[63] Liisa Postareff and Sari Lindblom-Ylänne. 2011. Emotions and confidence within teaching in higher education. *Studies in Higher education* 36, 7 (2011), 799–813.
[64] Mirko Raca and Pierre Dillenbourg. 2013. System for assessing classroom attention. In *Proceedings of the Third International Conference on Learning Analytics and Knowledge*. 265–269.
[65] Mirko Raca, Lukasz Kidzinski, and Pierre Dillenbourg. 2015. Translating head motion into attention-towards processing of student's body-language. In *Proceedings of the 8th international conference on educational data mining*.
[66] Nina Rajcic and Jon McCormack. 2020. Mirror Ritual: An Affective Interface for Emotional Self-Reflection. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. 1–13.
[67] Daniel C Richardson, Nicole K Griffin, Lara Zaki, Auburn Stephenson, Jiachen Yan, Thomas Curry, Richard Noble, John Hogan, Jeremy I Skipper, and Joseph T Devlin. 2020. Engagement in video and audio narratives: contrasting self-report and physiological measures. *Scientific Reports* 10, 1 (2020), 1–8.
[68] Verónica Rivera-Pelayo, Johannes Munk, Valentin Zacharias, and Simone Braun. 2013. Live interest meter: learning from quantified feedback in mass lectures. In *Proceedings of the Third International Conference on Learning Analytics and Knowledge*. 23–27.
[69] Paul Rozin and Adam B Cohen. 2003. High frequency of facial expressions corresponding to confusion, concentration, and worry in an analysis of naturally occurring facial expressions of Americans. *Emotion* 3, 1 (2003), 68.
[70] David C Rubin and Jennifer M Talarico. 2009. A comparison of dimensional models of emotion: Evidence from emotions, prototypical events, autobiographical memories, and words. *Memory* 17, 8 (2009), 802–808.
[71] James A Russell. 1980. A circumplex model of affect. *Journal of personality and social psychology* 39, 6 (1980), 1161.
[72] Philipp Schmidt, Felix Biessmann, and Timm Teubner. 2020. Transparency and trust in artificial intelligence systems. *Journal of Decision Systems* 29, 4 (2020), 260–278.
[73] Hong Shen, Haojian Jin, Ángel Alexander Cabrera, Adam Perer, Haiyi Zhu, and Jason I Hong. 2020. Designing Alternative Representations of Confusion Matrices to Support Non-Expert Public Understanding of Algorithm Performance. *Proceedings of the ACM on Human-Computer Interaction* 4, CSCW2 (2020), 1–22.
[74] David J Shernoff and Mihaly Csikszentmihalyi. 2009. Cultivating engaged learners and optimal learning environments. *Handbook of positive psychology in schools* 131 (2009), 145.
[75] David J Shernoff, Mihaly Csikszentmihalyi, Barbara Schneider, and Elisa Steele Shernoff. 2014. Student engagement in high school classrooms from the perspective of flow theory. In *Applications of flow in human development and education*. Springer, 475–494.
[76] Chaklam Silpasuwanchai, Xiaojuan Ma, Hiroaki Shigemasu, and Xiangshi Ren. 2016. Developing a comprehensive engagement framework of gamification

for reflective learning. In *Proceedings of the 2016 ACM Conference on Designing Interactive Systems*. 459–472.
[77] David A Slykhuis, Eric N Wiebe, and Len A Annetta. 2005. Eye-tracking students' attention to PowerPoint photographs in a science education setting. *Journal of Science Education and Technology* 14, 5 (2005), 509–520.
[78] Ömer Sümer, Patricia Goldberg, Sidney D'Mello, Peter Gerjets, Ulrich Trautwein, and Enkelejda Kasneci. 2021. Multimodal engagement analysis from facial videos in the classroom. *arXiv preprint arXiv:2101.04215* (2021).
[79] Wei Sun, Yunzhi Li, Feng Tian, Xiangmin Fan, and Hongan Wang. 2019. How Presenters Perceive and React to Audience Flow Prediction In-situ: An Explorative Study of Live Online Lectures. *Proceedings of the ACM on Human-Computer Interaction* 3, CSCW (2019), 1–19.
[80] Zhida Sun, Sitong Wang, Wenjie Yang, Onur Yürüten, Chuhan Shi, and Xiaojuan Ma. 2020. " A Postcard from Your Food Journey in the Past" Promoting Self-Reflection on Social Food Posting. In *Proceedings of the 2020 ACM Designing Interactive Systems Conference*. 1819–1832.
[81] Shamik Sural, Gang Qian, and Sakti Pramanik. 2002. Segmentation and histogram generation using the HSV color space for image retrieval. In *Proceedings. International Conference on Image Processing*, Vol. 2. IEEE, II–II.
[82] Daniel Szafir and Bilge Mutlu. 2012. Pay attention! Designing adaptive agents that monitor and improve user engagement. In *Proceedings of the SIGCHI conference on human factors in computing systems*. 11–20.
[83] Daniel Szafir and Bilge Mutlu. 2013. ARTFul: adaptive review technology for flipped learning. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. 1001–1010.
[84] Wenzhao Tan and Gang Rong. 2003. A real-time head nod and shake detector using HMMs. *Expert Systems with Applications* 25, 3 (2003), 461–466.
[85] Jaime Teevan, Daniel Liebling, Ann Paradiso, Carlos Garcia Jurado Suarez, Curtis von Veh, and Darren Gehring. 2012. Displaying mobile feedback during a presentation. In *Proceedings of the 14th international conference on Human-computer interaction with mobile devices and services*. 379–382.
[86] Van Thong Huynh, Soo-Hyung Kim, Guee-Sang Lee, and Hyung-Jeong Yang. 2019. Engagement Intensity Prediction with Facial Behavior Features. In *2019 International Conference on Multimodal Interaction*. 567–571.
[87] Milka Trajkova and Francesco Cafaro. 2018. Takes Tutu to ballet: Designing visual and verbal feedback for augmented mirrors. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 2, 1 (2018), 1–30.
[88] UNESCO. 2021. Education: From disruption to recovery. https://en.unesco.org/covid19/educationresponse. Accessed September 9, 2021.
[89] Robert Williams and Roman Yampolskiy. 2021. Understanding and Avoiding AI Failures: A Practical Guide. *Philosophies* 6, 3 (2021), 53.
[90] Rainer Winkler, Sebastian Hobert, Antti Salovaara, Matthias Söllner, and Jan Marco Leimeister. 2020. Sara, the lecturer: Improving learning in online education with a scaffolding-based conversational agent. In *Proceedings of the 2020 CHI conference on human factors in computing systems*. 1–14.
[91] Alexandra Witze. 2020. Universities will never be the same after the coronavirus crisis. *Nature* 582, 7811 (2020), 162–165.
[92] Jianming Wu, Bo Yang, Yanan Wang, and Gen Hattori. 2020. Advanced Multi-Instance Learning Method with Multi-features Engineering and Conservative Optimization for Engagement Intensity Prediction. In *Proceedings of the 2020 International Conference on Multimodal Interaction*. 777–783.
[93] Xiang Xiao and Jingtao Wang. 2015. Towards attentive, bi-directional MOOC learning on mobile devices. In *Proceedings of the 2015 ACM on International Conference on Multimodal Interaction*. ACM, 163–170.
[94] Xiang Xiao and Jingtao Wang. 2017. Undertanding and detecting divided attention in mobile mooc learning. In *Proceedings of the 2017 CHI conference on human factors in computing systems*. 2411–2415.
[95] Yukang Yan, Chun Yu, Wengrui Zheng, Ruining Tang, Xuhai Xu, and Yuanchun Shi. 2020. FrownOnError: Interrupting Responses from Smart Speakers by Facial Expressions. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. 1–14.
[96] Bin Yang and Cheng Huang. 2021. Turn crisis into opportunity in response to COVID-19: experience from a Chinese University and future prospects. *Studies in Higher Education* 46, 1 (2021), 121–132.
[97] Fang-Ying Yang, Chun-Yen Chang, Wan-Ru Chien, Yu-Ta Chien, and Yuen-Hsien Tseng. 2013. Tracking learners' visual attention during a multimedia presentation in a real classroom. *Computers & Education* 62 (2013), 208–220.
[98] Matin Yarmand, Jaemarie Solyst, Scott Klemmer, and Nadir Weibel. 2021. "It Feels Like I am Talking into a Void": Understanding Interaction Gaps in Synchronous Online Classrooms. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. 1–9.
[99] Telmo Zarraonandia, Paloma Díaz, Álvaro Montero, Ignacio Aedo, and Teresa Onorati. 2019. Using a Google Glass-based Classroom Feedback System to improve students to teacher communication. *IEEE Access* 7 (2019), 16837–16846.
[100] Haipeng Zeng, Xinhuan Shu, Yanbang Wang, Yong Wang, Liguo Zhang, Ting-Chuen Pong, and Huamin Qu. 2020. EmotionCues: Emotion-oriented visual summarization of classroom videos. *IEEE transactions on visualization and computer graphics* 27, 7 (2020), 3168–3181.
[101] Xucong Zhang, Seonwook Park, Thabo Beeler, Derek Bradley, Siyu Tang, and Otmar Hilliges. 2020. ETH-XGaze: A large scale dataset for gaze estimation under extreme head pose and gaze variation. In *European Conference on Computer Vision*. Springer, 365–381.
[102] Xucong Zhang, Yusuke Sugano, and Andreas Bulling. 2019. Evaluation of appearance-based methods and implications for gaze-based applications. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. 1–13.
[103] Xucong Zhang, Yusuke Sugano, Andreas Bulling, and Otmar Hilliges. 2020. Learning-based region selection for end-to-end gaze estimation. In *British Machine Vision Conference (BMVC 2020)*.
[104] Yunfeng Zhang, Q Vera Liao, and Rachel KE Bellamy. 2020. Effect of confidence and explanation on accuracy and trust calibration in AI-assisted decision making. In *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency*. 295–305.
[105] Zhengyou Zhang. 2012. Microsoft kinect sensor and its effect. *IEEE multimedia* 19, 2 (2012), 4–10.
[106] Bin Zhu, Xinjie Lan, Xin Guo, Kenneth E Barner, and Charles Boncelet. 2020. Multi-rate attention based gru model for engagement prediction. In *Proceedings of the 2020 International Conference on Multimodal Interaction*. 841–848.
[107] Sacha Zyto, David Karger, Mark Ackerman, and Sanjoy Mahajan. 2012. Successful classroom deployment of a social document annotation system. In *Proceedings of the sigchi conference on human factors in computing systems*. 1883–1892.

APPENDICES

A INSTRUCTOR PREFERENCE IN STUDENT LEARNING STATUS

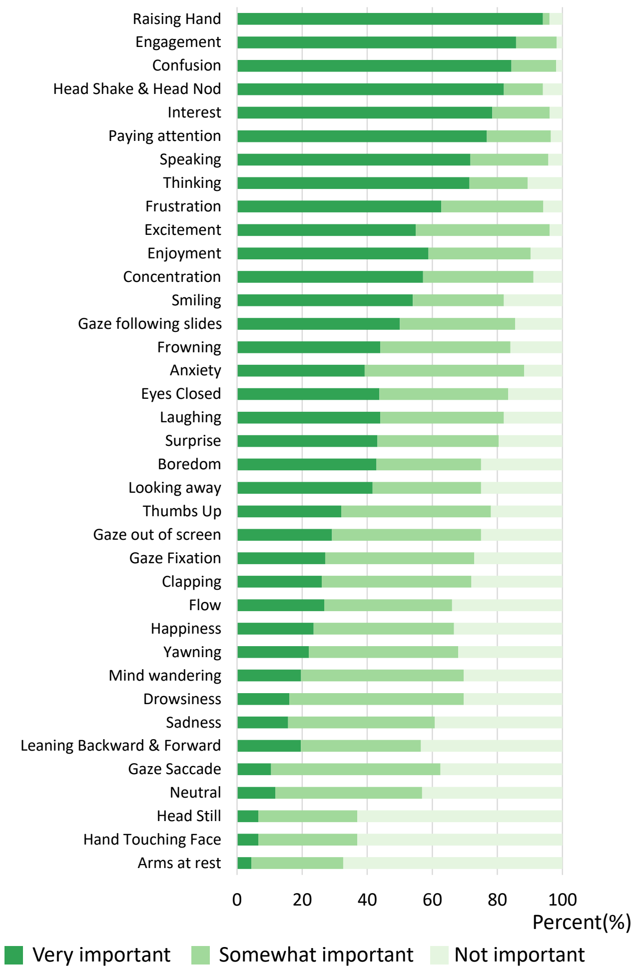


Figure 11: Instructors' preferences for students' learning status in synchronous online classes ranked in descending order. For each behavior/state, instructors gave their perceived importance in a 3 point scale, *Very important, Somewhat important, Not important*. From the figure, we can see that even for a top-ranked state/behavior, there are some instructors perceiving it as not important. On the contrary, even for a bottom-ranked state/behavior, there are some instructors perceiving it as very important. Hence, different individual preferences exist among instructors.

B STUDENTS' PERCEPTIONS OF THE USE OF *GLANCEE* IN PRACTICAL SYNCHRONOUS ONLINE CLASSES

From students' feedback in the 7-point Likert post-task questionnaire, we found that students did not feel nervous (M=2.25, SD=0.14), anxious (M=2.26, SD=0.15), uncomfortable (M=2.1, SD=0.13), or being distracted by the student end of *Glancee* (M=2.3, SD=0.14). And being monitored only made them somewhat engaged (M=4.21, SD=0.21). Overall, students were willing to use such a tool in the future (M=5.36, SD=0.15).

During the study, we found that no student participants turned on the camera (several students turned on the camera and then turned it off because others didn't turn it on) due to the reasons of discomfort and privacy concerns which are similar with the feedback we obtained from the exploratory survey (Sec. 3) and in line with [98]. From the interviews, we found that 94.4% of students did not worry about the privacy issue after knowing how the system worked. And 88.7% of students did not feel nervous, 96.3% of students did not feel anxious, 90.6% of students did not feel any uncomfortable due to the "anonymity protection". In terms of being distracted, 94.3% of students did not feel being distracted by the system, and some students even "*forgot there was a system running on my PC*." In addition, 45% of students felt more regulated because of "*being monitored*". And 48% of students felt their engagement was improved because of "*feeling being supervised*". The rest students did not feel more engaged because they thought "*the system was not a strong supervision as the data was anonymous*". In terms of their acceptance of the system's future adoption, 94.3% students would accept this system, because i) the privacy is protected, ii) students thought providing learning status to teachers could "*help teachers adjust the teaching*" which is both "*beneficial for students*" and "*a kind of respect for teachers*", and iii) the system will play a supervisory role for some students.

Overall, from the open-ended feedback, we found students appreciated the system in synchronous online classes. For example, S22 (Male, age: 22) said "*I thought the teacher could 'feel' me. When I expressed my confusion, the teacher explained the question in time, just like the teacher was paying attention to me.*" S27 (Female, age: 21) said "*In the previous online class, I was shy to ask questions. Now I can "ask a question" anonymously by showing a confused expression.*"

Finally, when asked for suggestions of the system, they gave some valuable feedback. First, the system should be made into a zoom plug-in to allow students to control whether it is turned on or not. Second, this tool should not take up too much memory and computing resources. Third, the accuracy of the detection algorithm should be guaranteed, otherwise the teacher will mistakenly understand the state of students. In addition, some students worries about "*gaming the system*", and some students would like to see the detected learning status of themselves.

C DESIGNED QUESTIONS IN QUESTIONNAIRES AND INTERVIEWS

The following describes the detailed questions in in-task questionnaires, post-task questionnaires, and interviews.

Table 5: In-task questionnaire for instructors in 7-point Likert scale (1: Not at all, 7: Very Much)

Category	Question
Usability of the system	1) How easy to use was the platform? 2) How much do you feel the platform helped you deliver the lecture? 3) How distracting was the platform when delivering the lecture? 4) How satisfied are you with the platform? 5) How much would you like to give lecture with the platform in the future?
Effectiveness of the system	1) Does the platform help you reflect upon and adjust the lecture? 2) Is it easy for you to find and locate the points which deserve recall and reflection? 3) How much connections did you feel with the students? 4) How aware were you of your lecture performance? 5) How easy was to observe students' reactions and learning status? 6) How easy was to respond to students' reactions and learning status? 7) How did you feel the sense of control of the class? 8) How satisfied are you with the lecture you gave? 9) How engaging was the lecture? 10) What is the overall quality of the lecture?
Effects on instructors' emotions	1) Did you feel engaged? 2) Did you feel excited 3) Did you feel exhausted? 4) Did you feel upset? 5) Did you feel happy? 6) Did you feel hopeful? 7) Did you feel overwhelmed? 8) Did you feel sense of safety? 9) Did you feel nervous? 10) Did you feel anxious? 11) Did you feel confident?
Effects on instructor' workload	1) How is your cognitive load? 2) How is your attentional load? 3) How is your workload? 4) How much time pressure did you feel? 5) How much time do you spend on observing students' reactions and learning status?

Table 6: Post-task questionnaire for instructors in 7-point Likert scale (1: Not at all, 7: Very Much)

Category	Question
Perceptions of the usefulness of each proposed feature	1) How do you think of the usefulness of triggered behaviors (head/facial behaviors)? 2) How do you think of the usefulness of engagement? 3) How do you think of the usefulness of confusion? 4) How do you think of the usefulness of gaze concentration? 5) How do you think of the usefulness of emotion? 6) How do you think of the usefulness of Adaptable Chart Type? 7) How do you think of the usefulness of Adaptable Chart Color? 8) How do you think of the usefulness of Adaptable Hidden/Display of views? 9) How do you think of the usefulness of Adaptable Chart Position and Size? 10) How do you think of the usefulness of Reminder? 11) How do you think of the usefulness of Adaptable Reminder Mechanism? 12) How do you think of the usefulness of Adaptable Reminder Threshold? 13) How do you think of the usefulness of Design of Sidebar? 14) How do you think of the usefulness of Post-class View? 15) How do you think of the usefulness of historical data aligned with Slides Number? 16) How do you think of the usefulness of Adaptable Trigger?
Trust	17) What is your perceived accuracy of the system? 18) What is your overall trust on the system? 19) What is you trust level on trigger behaviors? 20) What is your trust level on engagement? 21) What is your trust level on confusion? 22) What is your trust level on gaze concentration? 23) What is your trust level on emotion?

Table 7: Interview with instructors

Category	Question
Pattern and reasons of instructors' behaviors	1) When would you refer to the sidebar? 2) Why did you perform the behavior? Why? (The behavior should be specified with the instructors' actual behaviors during the teaching process) 3) How did you react to the situation? (The specific situation should be asked based on the actual situation occurred in the teaching process) 4) Have you ever adjusted the teaching according to students' reactions? Could you give an example and explain why? 5) Why did you turn off the view? / Why did you not turn of any view? 6) How did you deal with the obvious decline in the state of students? 7) Why did you configure the system like this (the rest part of the question should be asked based on instructors' actual configurations in the system, e.g., change to gauge chart/set the reminder to always remind)?
Effects on instructors	8) Did the system affect your emotions? (Could you explain a little bit more?) 9) Did you feel the system distract you? Why? 10) How about your cognitive load? When did you feel the cognitive load? Why? 11) How about your attentional load? When did you feel the attentional load? Why? 12) How about your workload? When did you feel the workload? Why?
User preference	13) What view (What student learning status) did you care about during the class? Why?
Trust	14) Did you trust the system? Why? 15) Among the learning states displayed in the system, which do you believe more? Why? 16) Have you ever doubted the system? When? Why?
Usability and usefulness	17) How do you think of the usability and usefulness of the system? 18) Do you think this system has affected the quality of your classes? 19) Do you think this system helps you establish a better connection with your students? Why? Could you provide some examples? 20) Which parts do you find useful in the system? 21) How do you think of the reminder function? (The reminder here can be replaced with other features, according to instructors' actual interaction history) Why? 22) What do you think are the advantages and disadvantages of our system compared with *EngageClass* (the system with only engagement displayed)? 23) What do you think are the advantages and disadvantages of our system compared with *ZoomOnly*? 24) What do you think of this system compared with the zoom with only some students turning on the camera? 25) How do you think this system compares to the zoom with all the students turning on the camera?
Further thoughts, suggestions, and comments	26) What is your overall feelings and perceptions? 27) Could you give some suggestions about the system? 28) How do you think of the future online classes? 29) Do you think students should see this system? Why? 30) Do you think it's better to show students' status through sidebar on one screen or on the second screen?

Table 8: Post-task questionnaire for students in 7-point Likert scale (1: Not at all, 7: Very Much)

Question
1) Did you feel nervous during the class?
2) Did you feel anxious during the class?
3) Did you feel distracted by the monitoring system?
4) Did you feel uncomfortable?
5) Did you think the system make you more engaged in the class?
6) How much would you like to use this system in the future?

Table 9: Interview with students

Category	Question
Perceptions	1) I observed that you turned on/did not turn on the camera in Zoom, why? 2) How was your overall feelings during the class? 3) How did you feel when you were being monitored by the system? 4) What is your concerns about the privacy? 5) (Repeat how our system works) In this case, do you worry about the privacy issue? 6) Did you feel nervous during the class? Why? 7) Did you feel anxious during the class? Why? 8) Did you feel being distracted by the monitoring system? Why? 9) Did you feel uncomfortable when being monitored by the system? Why? 10) Did you feel more engaged (compared with no monitoring)? Why?
Acceptance and suggestions	11) Would you like to accept such a system? Why? 12) What do you think of the real adoption of such a system in the future online classroom? 13) What improvements do you think the system needs to make for future adoption? 14) Do you have any other suggestions or comments?