Investigating the Effects of a Real-time Student Monitoring Interface on Instructors' Monitoring Practices in Online Teaching

Hayeon Lee*
hci+d lab.
Seoul National University
Seoul, South Korea
raon0172@snu.ac.kr

Seora Park*
hci+d lab.
Seoul National University
Seoul, South Korea
annieseora96@snu.ac.kr

Esther Hehsun Kim*
hci+d lab.
Seoul National University
Seoul, South Korea
ehk@snu.ac.kr

Jiyeon Seo
hci+d lab.
Seoul National University
Seoul, South Korea
jy1126@snu.ac.kr

Hajin Lim
hci+d lab.
Seoul National University
Seoul, South Korea
hajin@snu.ac.kr

Joonhwan Lee
hci+d lab.
Seoul National University
Seoul, South Korea
joonhwan@snu.ac.kr

ABSTRACT

The shift to online education, accelerated by the COVID-19 pandemic, has introduced challenges in monitoring student engagement, an essential aspect of effective teaching. In response, real-time student monitoring interfaces have emerged as potential tools to aid instructors, yet their efficacy has not been thoroughly examined. Addressing this gap, we conducted a controlled experiment with 20 instructors examining the impact of engagement cues (presence versus absence) and student engagement levels (high versus low) on instructors' monitoring effectiveness, teaching behavior adjustments, and cognitive load in online classes. Our findings underscored the fundamental benefits of student engagement monitoring interfaces for improving monitoring quality and effectiveness. Furthermore, our study highlighted the critical need for customizable interfaces that could balance the informational utility of engagement cues with the associated cognitive load and psychological stress on instructors. These insights may offer design implications for the design of future student engagement monitoring interfaces.

CCS CONCEPTS

• **Human-centered computing** → *Empirical studies in HCI*; *Empirical studies in visualization.*

KEYWORDS

Real-time student monitoring system, monitoring, student engagement, online teaching, online learning

ACM Reference Format:
Hayeon Lee, Seora Park, Esther Hehsun Kim, Jiyeon Seo, Hajin Lim, and Joonhwan Lee. 2024. Investigating the Effects of a Real-time Student Monitoring Interface on Instructors' Monitoring Practices in Online Teaching. In *Proceedings of the CHI Conference on Human Factors in Computing Systems (CHI '24), May 11–16, 2024, Honolulu, HI, USA.* ACM, New York, NY, USA, 11 pages. https://doi.org/10.1145/3613904.3642845

*These authors equally contributed to this research.

CHI '24, May 11–16, 2024, Honolulu, HI, USA

ACM ISBN 979-8-4007-0330-0/24/05
https://doi.org/10.1145/3613904.3642845

1 INTRODUCTION

The COVID-19 pandemic has expedited the transition to online classes, providing opportunities for quality education such as data-driven intervention [4, 32]. However, remote education, particularly in real-time online classes, poses challenges for instructors. These challenges arise from the limitations of visual and communication cues in capturing students' engagement during online classes [7]. Various cues available in traditional classroom settings, such as facial expressions and gestures, which are crucial for monitoring in-class behavior and learning activities, have significantly diminished in online education environments. Recent research indicates that instructors teaching online classes struggle to closely observe or actively interact with their students, thereby limiting their ability to monitor student behavior and employ effective teaching strategies based on student engagement [7, 23, 24, 26, 29, 30].

Monitoring student behavior is foundational to effective teaching, enabling instructors to gauge students' progress and manage classes efficiently [2, 3, 26]. It also facilitates understanding of individual learners' needs, allowing for timely interventions [6, 13, 20, 25], and helps anticipate and prevent potential issues like dropouts [9]. Moreover, monitoring fosters stronger instructor-student relationships [26]. Student monitoring remains essential in online teaching, too. In online education, where concentration and technological issues can lead to decreased student engagement [16], continuous monitoring of student engagement levels is crucial [1, 13, 26], especially given the limited opportunities for instructor-student interaction [5, 8, 33]. However, real-time student engagement monitoring in online classes presents challenges, hindering personalized instruction and resulting in reduced student engagement [7, 16, 33].

To address these challenges, researchers have been actively designing and evaluating real-time student monitoring interfaces for online classes [14, 19, 21, 31]. For instance, Glancee provides comprehensive real-time visualizations of students' learning metrics [17]. Further, AffectiveSpotlight utilizes facial expression and head gesture analysis to identify engaged audiences to reduce presenters' anxiety [22]. Similarly, Groupnamics facilitates group discussions in online classes by visualizing the status of group discussions, thereby facilitating a deeper understanding of group dynamics [28].

As such, previous research has introduced various interfaces aimed at supporting instructors' monitoring efforts. These interfaces have offered a wide array of monitoring metrics and visualization methods, with two main components typically included: "displaying individual student engagement" and "representing overall class engagement." While prior studies have acknowledged the potential of student engagement monitoring interfaces, their focus has primarily been on system development and usability assessment. Evaluation metrics such as perceived usefulness, ease of use, and system satisfaction have been prioritized. However, this evaluation approach often lacks a thorough understanding of how engagement cues directly assist instructors in monitoring and adjusting their online teaching methods. Therefore, this paper aims to rigorously examine the impact of student engagement cues on instructors' monitoring capabilities and teaching adjustments.

To comprehensively assess the impact of student monitoring interfaces on instructors' student monitoring and teaching strategies, we conducted a controlled experiment involving twenty experienced instructors who taught real-time online classes at the university level. In this experiment, instructors were asked to teach 10-minute classes twice with or without student engagement cues. For the experimental setup, we utilized a simplified interface that highlighted both individual student and class-level engagement in intuitive ways, concentrating on the essential elements of student monitoring interfaces. Our emphasis on system simplicity aimed to untangle the complexities of monitoring metrics, visualization, and interface designs from the core effects of engagement cues. Furthermore, rather than conducting a field experiment, which could be susceptible to disruptions and unpredictable circumstances, we opted to use pre-recorded videos of students displaying both engaged and disengaged behaviors. This approach allowed us to explore how the overall class engagement level (high vs. low) interacted with the presence of student engagement cues (present vs. absent).

Overall, our findings suggest that the interface displaying engagement cues significantly improved the accuracy of class monitoring and enhanced the perceived quality and quantity of monitoring. However, this improvement did not lead instructors to spend more time monitoring students. In addition, participants reported that the presence of engagement cues prompted them to make more teaching adjustments during the classes. Interestingly, the mere presence of engagement cues did not increase instructors' cognitive load. However, we observed an interaction effect where instructors' cognitive load was highest when engagement cues were present, but the ratio of highly engaged students was low. Based on these findings, we discuss the design implications of monitoring assistance systems in online teaching. Through post-task interviews, we delved into the reasons behind the quantitative findings and explored instructors' suggestions for improving student engagement interfaces. Based on these findings, we discuss the design implications of monitoring assistance systems in online teaching.

Our contribution is three-fold:

- Our controlled experiment design enabled us to rigorously investigate how the presence of student engagement cues (presence vs. absence) and student engagement levels (high vs. low) impacted instructors' monitoring quality, quantity, adaptive teaching behavior, and cognitive load in online classes. To disentangle the variations in diverse student engagement monitoring interfaces regarding monitoring metrics and visualization methods, we employed a probe system with a simplified interface containing only the core elements of diverse student engagement monitoring systems (individual- and class-level engagement visualization). Additionally, we utilized pre-recorded video of students to manage variations in student engagement levels.
- By incorporating both behavioral measures (quantifiable monitoring accuracy and monitoring quantity through eye-tracking) and subjective measures (perceived monitoring quality, quantity, and cognitive load), our findings provided comprehensive empirical evidence of the essential role that student engagement interfaces played in enhancing online teaching. Specifically, we discovered that simple engagement cues were effective in improving both the quantity and quality of monitoring, as well as in facilitating adjustments in teaching behavior. Additionally, our study identified an interaction between engagement cues and student engagement levels regarding cognitive load.
- Based on our findings, we offered design implications for future student monitoring systems for online teaching.

2 RESEARCH BACKGROUND

Active monitoring of student behavior is crucial for effective teaching, allowing instructors to assess progress and manage classes efficiently [2, 3, 26]. Monitoring helps understand individual learners' needs, enabling timely interventions [6, 13, 20, 25], and prevents issues like dropouts [9]. Given the close link between student engagement and academic achievement, instructors must actively assess student engagement to tailor teaching strategies and provide timely interventions [2].

However, instructors of online classes often face challenges that hinder their ability to monitor and sustain student engagement effectively. A key issue is the limited opportunities for real-time interactions and immediate feedback in virtual classrooms [26]. Consequently, instructors must navigate the complexities of monitoring and maintaining student attention and participation without the benefits of physical presence and direct, non-verbal cues [5]. This inadequate monitoring can significantly diminish student engagement, given that the immediacy and dynamism of traditional classroom settings are frequently absent in online teaching environments.

Recent research has addressed these challenges by developing various systems aimed at effectively monitoring and visualizing students' engagement to support instructors in their monitoring efforts. Many studies have utilized learning analytics and computer vision techniques to capture students' non-verbal cues, including eye gaze, head movements, and facial expressions, enabling presenters and facilitators to discern levels of engagement [2, 11, 12, 17, 18, 22, 28, 30, 34]. Furthermore, these studies have introduced diverse visualization methods to illustrate behavioral, cognitive, and emotional engagement, as well as learning progress [2, 17] such as interactive dashboards [19] and spotlights interface [22].

As such, a rich body of research has introduced a range of interfaces designed to assist instructors in monitoring student engagement, typically comprising two key elements: displaying individual student engagement and representing overall class engagement (e.g., [28]). While these interfaces vary in monitoring metrics and visualization methods, they have generally received positive evaluations. However, existing studies have primarily focused on assessing the usability of their novel interfaces, thereby limiting the fundamental understanding of how engagement cues influence instructors' monitoring and teaching practices. Therefore, this study seeks to rigorously examine how the presence of student engagement cues affects instructors' ability to monitor students through a controlled laboratory experiment.

Given the challenges of observing student engagement in online environments due to the lack of non-verbal cues and limited interaction, we hypothesize that integrating engagement cues into the video conferencing interface for teaching would help instructors better gauge student engagement levels and adapt teaching methods accordingly. To examine the fundamental impact of student engagement on instructors' monitoring and teaching adjustments, we propose the following hypothesis:

- **H1**. The presence of student engagement cues, compared to their absence, would lead to increased (a) monitoring accuracy, (b) monitoring activity, and (c) teaching adjustment by instructors.

While previous studies have emphasized the beneficial role of engagement cues in facilitating monitoring for instructors, they have also underscored its potential to overwhelm instructors by increasing cognitive burden due to the need to process additional information during classes [28]. For instance, Ma et al. [17] observed an increased workload for instructors using the interface to monitor students' statuses in class. These effects on cognitive load may arise from the multitude of functionalities and complexity of visualizations of monitoring metrics on each interface. Thus, we aim to investigate the impact of student engagement cues presented in a simplified form on instructors' cognitive load, leading us to pose the following research question:

- **RQ1**. How does the presence of student engagement cues affect instructors' cognitive load when teaching online classes?

An important consideration is the level of student engagement in the class, as it can significantly impact instructors. For example, low engagement levels among students may lead to increased cognitive load for instructors, resulting in more frequent monitoring and attempts to adjust teaching methods [17]. To examine the impact of student engagement levels on instructors, we propose the following hypothesis:

- **H2**. A low level of class engagement, compared to high class will result in increased (a) monitoring activity, (b) teaching adjustment, and (c) cognitive load.

Moreover, it would be essential to explore how student engagement levels interact with the presence of engagement cues. While the presence of engagement cues can positively affect instructors, it may also increase cognitive load, especially when student engagement levels are low. To investigate this interaction, we pose the following research question:

- **RQ2**. How does the interaction between the presence of student engagement cues (present vs. absent) and the level of student engagement (high vs. low) influence instructors' monitoring accuracy, monitoring activity, teaching adjustment, and cognitive load in online classes?

3 METHOD

3.1 Study Design and Interface

Our experiment employed a 2x2 within-subjects design, manipulating the presence of engagement cues (absent vs. present) and student engagement levels (high vs. low). Participants were randomly assigned to two of the four experimental conditions (Figure 1) and instructed to run two 10-minute classes under different conditions (combinations (a)-(d) or (b)-(c) in random order).

3.1.1 Design of Student Engagement Cues. In manipulating the engagement cue, we utilized two interfaces: one without engagement cues (Figure 1 (a) and (b)) and one with engagement cues (Figure 1 (c) and (d)). These interfaces closely resembled the Zoom platform, commonly used in online classes, to facilitate the evaluation of student engagement cues in familiar settings. The interface comprised two main components: a screen-sharing section (left) for instructors to display class materials and a student window (right) showcasing 18 student videos in a gallery view.

The interface featuring engagement cues (Figure 1 (c) and (d)) provided engagement indicators in gallery view in two modes: 1) individual student engagement and 2) overall class engagement, aligning with the primary focus of existing student monitoring interfaces in prior studies (e.g., [17, 28]). Individual student engagement cues were displayed in the lower right corner of each student's screen, utilizing distinct colors for enhanced clarity as suggested by previous research [2]. Specifically, green color indicated a highly engaged student, while red indicated a student with low engagement. Additionally, the overall class engagement was provided on the far right through a bar graph accompanied by a numerical value (e.g., 64%), indicating the percentage of students demonstrating high levels of engagement. Bar graphs were chosen for their effectiveness as a visualization method for real-time class monitoring [30], ensuring instructors had a clear and intuitive representation of class-level engagement.

3.1.2 Manipulation of Student Engagement Levels. To ensure the systematic manipulation of student engagement levels during the experiment, we decided to utilize pre-recorded videos of students rather than involve real-time participants. We recruited 18 volunteers from online communities, each asked to record two 15-minute videos: one demonstrating high engagement and the other low engagement in an online class. Detailed guidelines were provided to these volunteers for each condition, outlining factors such as eye gaze, facial expressions, and head movement, which were established indicators of engagement in previous studies [2, 18, 22].

After collecting videos from volunteers, researchers meticulously reviewed them to ensure compliance with the provided instructions.

Figure 1: Four experiment conditions of the study. (a) engagement cues absent x low class engagement level (b) engagement cues absent x high class engagement level (c) engagement cues present x low class engagement level (d) engagement cues present x high class engagement level

Two researchers independently assessed the volunteers' videos, categorizing them as either "high engagement" or "low engagement" based on predefined criteria while being blind to the conditions. All videos provided by volunteers were successfully classified, resulting in a total of 36 videos (18 high-engagement and 18 low-engagement videos), which were then used to create simulated class environments representing highly engaged classes versus low-engaged classes. To simulate classes with high engagement levels, we included at least 14 out of 18 high-engagement videos in the student gallery view, resulting in an overall engagement level of 78% in the class (student engagement level: high condition). Conversely, to represent low overall class engagement, no more than 8 out of 18 high-engagement videos were included in the student gallery view, resulting in an overall engagement level of 44% in the class (student engagement level: low condition).

3.2 Participants

We recruited 20 instructors (Table 1) for our experiment (average age = 31.7, SD age = 5.25, 50% female) by posting announcements on online university communities. Participants had varying levels of teaching experience, ranging from less than six months to over five years, and all had experience teaching synchronous online courses at universities with class sizes of 20 to 30 students. Each participant received 50,000 Korean Won (equivalent to $42 USD) for their participation. The study procedure took approximately two hours to complete.

3.3 Study Procedures

Once participants expressed their willingness to participate in the study and scheduled their experiment sessions, we asked them to prepare their own class materials for conducting two 10-minute online classes. This approach aimed to provide them with a comfortable and familiar environment while teaching online classes for the experiment.

Upon arriving at the lab, researchers introduced the study's objective and obtained consent from the participants. Next, a brief explanation of the interface was provided, highlighting its ability to display students' engagement levels in real-time using students' gaze, facial expressions, and body movements. Subsequently, participants completed a pre-survey to provide demographic details and information about their teaching experience.

Before starting 10-minute online classes, we clarified that students would actively participate in their classes in real-time, although their microphones would be muted, and direct interaction with them during the class would not be possible due to the system being in development. Following this, the participants went through an eye-tracking calibration process to monitor their gaze movement during the lecture.

Table 1: Participants' demographic and teaching experience

Participants	Gender	Age	Teaching Experience	Subject	Highest Education Level
P1	Female	27	5 years	English	Master's student
P2	Female	27	less than 6 months	Society and Culture	Master's student
P3	Female	35	6 months-2 years	Oceanography	PhD degree
P4	Female	34	6 months-2 years	East Asian Studies	PhD student
P5	Female	27	less than 6 months	Digital Data	Master's degree
P6	Male	39	2-5 years	Music	PhD student
P7	Female	31	less than 6 months	English	Master's student
P8	Female	38	more than 5 years	Arts Law, Policy	PhD degree
P9	Female	29	less than 6 months	Intro to Computing	Master's degree
P10	Male	27	less than 6 months	Mathematical Statistics	Bachelor's degree
P11	Male	33	more than 5 years	Programming	PhD student
P12	Female	27	less than 6 months	Humans and Technology	Master's student
P13	Male	28	less than 6 months	Humans and Computer	Master's student
P14	Male	28	less than 6 months	History	Master's student
P15	Male	27	less than 6 months	Computer language	Master's student
P16	Male	39	2-5 years	Computing	PhD student
P17	Male	46	more than 5 years	Music and Technology	PhD student
P18	Male	28	6 months-2 years	Programming	Master's student
P19	Female	33	6 months-2 years	HCI	PhD degree
P20	Male	31	less than 6 months	Business, Finance	Bachelor's degree

Upon arranging the experimental settings and randomly assigning participants to one of the four conditions, researchers initiated eye-tracking recording and asked participants to start their first 10-minute online lectures. After their first lecture, participants were asked to complete a post-survey. Following that, participants were asked to deliver a second 10-minute lecture under different experimental conditions and fill out a second post-survey.

After the experiment, we conducted semi-structured interviews to gather participants' reflections on their experiences with the interfaces. Participants were prompted to describe how engagement cues in the interface influenced their monitoring behaviors during the class compared to when such cues were absent. We also asked them for feedback on the interfaces for monitoring student engagement. Following the interviews, we disclosed that the student videos were pre-recorded and provided participants with the option to opt out if they desired. Participants were then compensated for their participation and left the lab.

3.4 Measures

To assess the hypotheses and investigate research questions, we utilized various behavioral and self-reported subjective scales to measure instructors' monitoring accuracy, monitoring activity, teaching adjustment, and cognitive load.

3.4.1 Monitoring Accuracy. After each lecture, participants were asked to indicate the number of students who were highly engaged in the class. These responses were converted into the overall class engagement percentages (e.g., if they answered 12 highly engaged students out of 18, it was calculated into 66.7% overall class engagement). These percentages were then compared with the predetermined threshold percentages for high (78%) and low (44%) student engagement level conditions. The absolute difference between participants' and thresholds was used to measure their monitoring accuracy, defined as **"monitoring error."** A lower monitoring error indicated higher monitoring accuracy. Additionally, participants self-reported the quality of their class monitoring using six 7-point Likert scale questions from [22], including *"I was able to grasp the overall atmosphere of the class"* and *"I was able to capture the reactions of various students at a glance."* These questions created a reliable scale (Cronbach's alpha = .963) of **"perceived monitoring quality."**

3.4.2 Monitoring Activity. To examine how the presence of engagement cues influenced participants' monitoring activities during the class, we conducted an analysis of eye-tracking data. The **"proportion of participants' gazes on the student window"** was computed based on the total duration of gazes on the interface. Additionally, participants self-reported their monitoring quantity using three 7-point Likert scale questions from [22], including statements like *"I tried to look at the student screen often during class,"* and *"I watched the student screen for a long time during class."* These questions were used to create a reliable scale (Cronbach's alpha = .899) to measure **"perceived monitoring quantity."**

3.4.3 Teaching Adjustment. Participants self-reported the extent to which they adapted their teaching based on student engagement. Five 7-point Likert scale items adapted from [22], including statements like *"The contents of my class were adjusted according to the atmosphere in the class"* and *"My voice, tone, and speed of class* *were adjusted based on student engagement."* These items formed a reliable scale (Cronbach's alpha = .902) to measure **"perceived teaching adjustment behaviors."**

3.4.4 *Cognitive Load.*

We assessed how the presence of student engagement cues affected participants' cognitive load while running a class using an adapted version of NASA-TLX [10]. This version included four 7-point Likert scales, such as *"I found running this class mentally demanding"* and *"I found running this class stressful."* These scales formed a reliable measure of **"cognitive load"** (Cronbach's alpha = .813).

3.5 Interview Questions

After participants completed two 10-minute lectures and post-task surveys, we conducted semi-structured interviews, aiming to understand the reasons behind participants' experiences corresponding to both the behavior and self-reported findings. Specifically, we asked participants to describe how their monitoring activity, quality, teaching adjustment, and cognitive load differed when the engagement cue was available compared to when it was not. Finally, we gathered their recommendations for future enhancements to student engagement monitoring systems.

3.6 Data Analysis

3.6.1 *Quantitative Data Analysis.*

The quantitative data from the eye-tracking data and post-survey collected during the lab experiment were analyzed using SPSS 25.0. To test the hypotheses and investigate the research question, we conducted Mixed Model ANOVAs with a 2 (engagement cues: present vs. absent) by 2 (student engagement level: high vs. low) design on all measures described above. As the study employed a within-subjects design, we examined the potential effects of trial orders using MANOVA for all measures. However, we did not observe any significant effect of the trial order. Consequently, we did not include the order factors in our analysis.

3.6.2 *Qualitative Data Analysis.*

We analyzed the data obtained from post-task interviews through a thematic coding approach [27]. After transcribing all post-task interviews, the first author segmented the interview transcripts into distinct topics, including *"monitoring quality (accuracy)," "monitoring quantity," "teaching adjustment"* and *"overall satisfaction and feedback."* Subsequently, all authors individually reviewed and coded the interview transcripts to identify common and interesting patterns. Following this initial coding phase, all authors shared their respective codes and collaboratively developed themes through an iterative process.

4 RESULTS

4.1 Monitoring Accuracy

We hypothesized in H1(a) that the presence of student engagement cues would result in increased monitoring accuracy and inquired in RQ2 about the interaction effect of the presence of engagement cues and the student engagement levels on monitoring accuracy. To test this hypothesis and RQ, we conducted two mixed-model ANOVAs, one using monitoring error and the other using perceived monitoring quality as the dependent measures.

We found that the monitoring error (Figure 2 (a)) was significantly lower when student engagement cues were present (M = 10.55, SE = 12.22) compared to when they were absent (M= 26.25, SE = 43.92) (F[1, 27.07] = 19.226, p<.001, $\eta 2$ = .41). Also, the results yielded that the perceived monitoring quality (Figure 2 (b)) was significantly higher when student engagement cues were present (M = 5.04, SE = 0.32) compared to when they were absent (M = 3.25, SE = 0.31) (F[1, 20.85] = 12.738, p = .002, $\eta 2$= .37). Thus, H1(a) was supported. However, there was no significant interaction effect between the presence of student engagement cues and the level of student engagement (p > .05) on monitoring accuracy (RQ2(a)).

During the post-task interviews, participants noted that the colored visualizations of student engagement levels and the overall engagement rate in the bar graph helped them better understand the overall atmosphere of the class. For example, P14 said, *"In the first class (condition using the interface without engagement cues), it was vague whether students understood me or not. However, in the second class (condition using the interface with engagement cues), it was definitely helpful as the green lights were on and off. It was good to see and check whether students were focusing on me or not in a short glance."* Additionally, the majority of participants mentioned that the improvement in overall class monitoring quality strengthened their sense of connection with students. They felt that they could recognize and react to their students much better with the student engagement cues and felt closer to them. For example, P3 stated, *"After two years of teaching students via ZOOM, at one point, it felt like my students didn't exist there anymore. But when I used the system, I felt more connected to them".*

4.2 Monitoring Activity

We hypothesized that monitoring activities would be increased when student engagement cues were present compared to when they were absent (H1(b)) and when student engagement levels were low (H2(a)). Also, we investigated the interaction effect between engagement cues and student engagement level on monitoring activity (RQ2). We tested these hypotheses and RQ by conducting two mixed-model ANOVAs, one using the proportion of gaze on the student window and the other using perceived monitoring quantity as the dependent measure.

In the analysis of the proportion of gaze on the student window (Figure 2 (c)), we did not find any significant effect of the presence of engagement cues (when engagement cues were present: M = 21, SE = 3.12; when engagement cues were absent: M = 22, SE = 2.56) (p > .05). Additionally, there was no significant effect of the student engagement level (p > .05) and no interaction effect between engagement cues and student engagement level (p > .05).

In contrast, participants perceived that they engaged in significantly more monitoring activities (Figure 2 (d)) when engagement cues were present (M = 5.3, SE = 0.35) compared to when engagement cues were absent (M = 4.97, SE = 0.34) (F[1, 22.56] = 5.463, p = .029, $\eta 2$ = .19). However, there was no significant effect of the student engagement level on perceived monitoring quantity (p > .05) and no interaction effect between engagement cues and student engagement level (p > .05).

We observed some clues regarding the reason for the discrepancy between their actual monitoring behavior (proportion of eye gaze

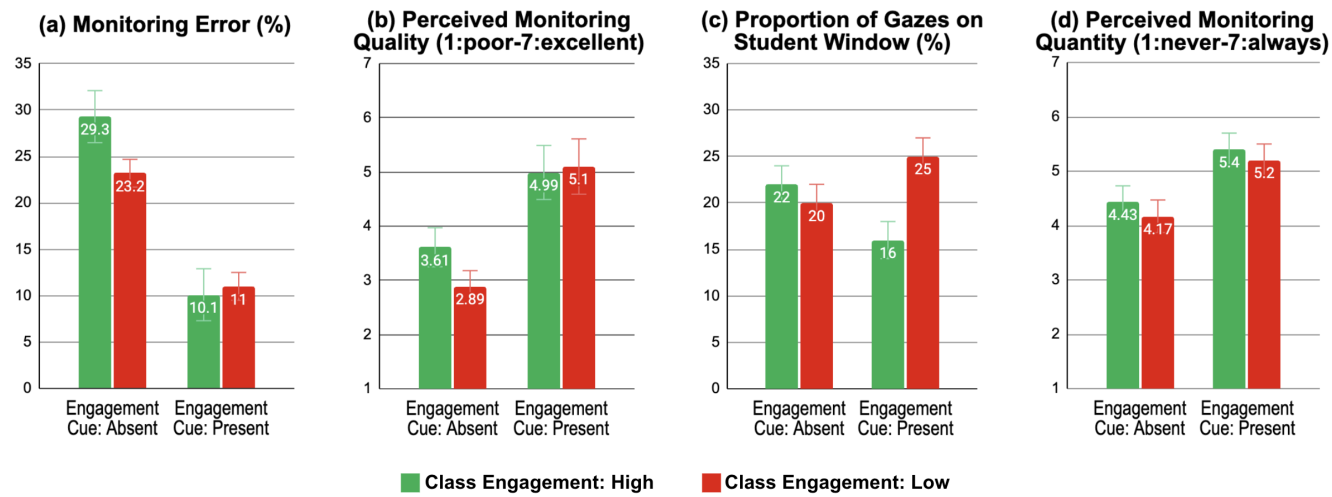

Figure 2: Effects of engagement cues and student engagement levels on monitoring accuracy ((a) monitoring error, (b) perceived monitoring quality) and monitoring activity ((c) the proportion of eye gazes on the student window, (d) perceived monitoring quantity).

on the student window) and their perceptions of engaging in monitoring activities (perceived monitoring quantity) during post-task interviews. Many participants mentioned that they were drawn to the student window more often primarily because of the color changes in the students' videos, which led to the perception that they had monitored students more frequently when engagement cues were present. For example, P11 said, *"I was almost forced to look at it when something turned red."* At the same time, they noted that when engagement cues were present, their monitoring became much quicker: *"It (the interface with engagement cues) was really good to know students' engagement level quickly at a short glance"* (P14). Consequently, this result indicated that their frequency of looking at the student window might have increased when engagement cues were present. However, they might have spent less time monitoring students overall, as they could quickly assess students' engagement levels.

4.3 Teaching Behavior Adjustment

We hypothesized that participants would report adjusting their teaching more frequently when student engagement cues were present compared to when they were absent (H1(c)) and when student engagement levels were low (H2(b)). Also, we investigated the interaction effect between engagement cues and student engagement level on teaching behavior adjustment (RQ2). We tested these hypotheses and RQ by conducting a mixed-model ANOVA using the measure of perceived teaching adjustment behaviors as the dependent variable.

As shown in Figure 3 (a), we found a significant effect on perceived teaching adjustment behaviors ($F[1, 20.92] = 4.189$, $p = .05$, $\eta 2 = .16$). Participants perceived that they adjusted their teaching styles based on student engagement more often when engagement cues were present ($M = 4.68$, $SE = 0.34$) compared to when engagement cues were absent ($M = 3.91$, $SE = 0.34$). However, there was no significant effect of student engagement ($p > .05$) and no interaction effect ($p > .05$) on perceived teaching adjustment behaviors.

The analysis of post-task interviews revealed findings similar to the quantitative results. Participants mentioned that they tried to adjust their teaching behaviors more often when engagement cues were present. While they could not interact directly with students in the experiment, participants reported that they implemented various strategies to engage the low-engagement students identified in the interface showing engagement cues. For example, some participants commented that they adjusted their voice volume, pitch, and speed when they noticed low student engagement: *"When the participation rate dropped, and lots of red lights turned on, I tried to slow down and give additional explanations"* (P14). Moreover, in order to get students' engagement, P5 said he asked questions, while P1 said she made a joke and introduced interesting examples. P18 perceived the low overall student engagement cues as a signal *"to move on quickly to another topic"* (P18) and cut out some parts of the lecture.

4.4 Cognitive Load

We inquired about how the presence of engagement cues affected the participants' cognitive load (RQ1) and whether the presence of engagement cues and student engagement levels had an interaction effect on the participants' cognitive load (RQ2). Also, we hypothesized that the low engagement level of students would increase the participants' cognitive load. We examined the hypothesis and RQs by conducting a mixed-model ANOVA using the measure of cognitive load as the dependent variable. We did not find a significant effect of the engagement cues ($p > .05$) or the student engagement level ($p > .05$) on cognitive load. Interestingly, there was a significant interaction effect between the engagement cues and the student engagement levels ($F[1, 20.88] = 4.440$, $p = .047$, $\eta 2 = .17$). As shown in Figure 3 (b), post-hoc analysis revealed that participants' cognitive load was highest when the engagement cues were present,

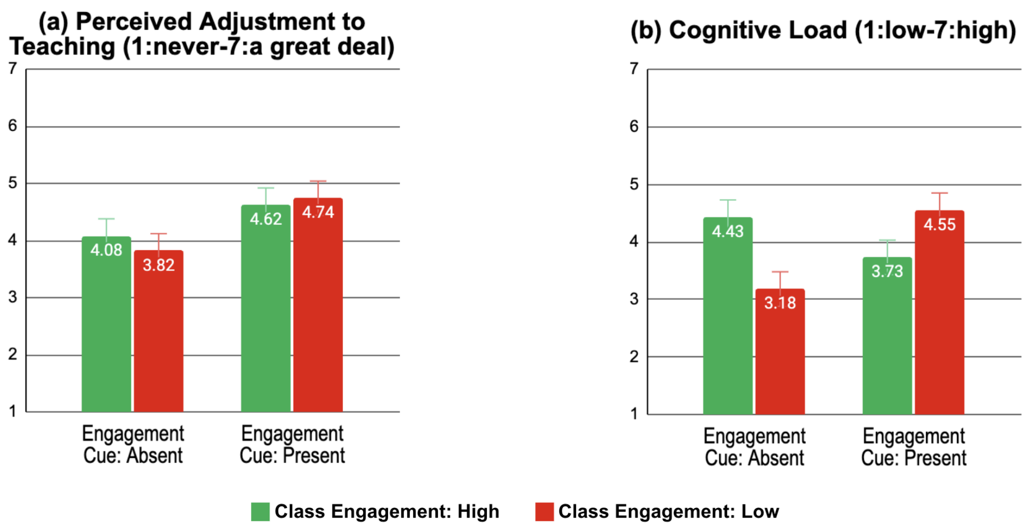

Figure 3: Effects of engagement cues and student engagement levels on perceived teaching adjustment and cognitive load

and the ratio of highly engaged students was low (M= 4.55, SE = 0.31). In contrast, participants' cognitive load was lowest when the engagement cues were absent, and the ratio of highly engaged students was low (M=3.17, SE = 0.32).

Post-task interviews revealed opposing views of the effect of the engagement cues on cognitive load. Some participants perceived that their cognitive load was reduced when the engagement cues were present in the interface, as they did not have to process the nonverbal expressions of individual students to gauge their engagement. In contrast, others found that engagement cues in the interface increased their cognitive load as they generally engaged in more monitoring activities: *"When there was no system, I could ignore some students who did not concentrate on me and my class. But with the system, I felt like I had to do something for those students."* (P12). These contrasting views might account for the insignificant effect of the engagement cues conditions on cognitive load.

Further, participants reported feeling pressured and even embarrassed when they observed low student engagement through the interface with engagement cues, which aligned with the interaction effect in the quantitative result. For example, P7 said, *"When the red light was on, it was a bit burdensome to look at the student screen. When the light was on, I felt like I was being evaluated in real-time, so I felt more pressure."* Such tensions and embarrassments tended to be higher among participants with shorter teaching careers, as P4 mentioned: *"There was psychological pressure. If all the red lights were on, novices like me would be more nervous and frustrated. Like, everyone is not concentrating! Then, what should I do?"*.

4.5 Suggestions for Future Student Engagement Monitoring Interfaces

At the end of the post-task interviews, participants provided diverse suggestions for enhancing future student engagement monitoring interfaces. One prevalent suggestion was the customization of a monitoring interface that would allow them to toggle the engagement visualization on and off as needed during class. In particular, P1 suggested that this feature would help manage observing students' reactions separately from delivering class content.

Additionally, many participants suggested displaying only one type of engagement (i.e., either showing only highly engaged or lowly engaged students). For instance, P5 preferred to see only red lights (disengaged students) as it would help identify students needing assistance quickly. Interestingly, instructors' preferences regarding which to highlight varied with their teaching experience. Instructors with more teaching experience preferred to showcase only students who were not engaging, as they believed that such students needed a greater amount of intervention and support. For instance, P16, an experienced teacher, suggested focusing on low engagement students rather than monitoring high engagement students as they were already engaged, requiring less monitoring. Conversely, participants with less teaching experience found visualizations of low engagement disheartening and stressful. For example, P2, with less than six months of experience, expressed feeling discouraged and unable to focus on the class when low levels of student engagement were displayed.

Further, P19 proposed presenting a score of the average student engagement accumulated over the course of each class to allow instructors to track the overall level of engagement of the class. This suggestion aimed to assist instructors in developing teaching plans for future classes and avoiding over-critiquing students who might be unmotivated or disengaged at a particular moment.

Finally, some instructors suggested alternative approaches to distinguishing student engagement levels. A proposal from Participant 13 suggested categorizing engagement levels on a high-average-low scale instead of a binary high-low scale to ease the standard for "low engagement." He said, this would reduce the burden of students being assessed constantly.

5 DISCUSSION

The purpose of this study was to rigorously investigate the effects of student engagement cues from the instructor's perspective. We systematically examined how the presence of student engagement cues, along with varying levels of student engagement, influenced instructors' monitoring quality, quantity, teaching behavior adjustment, and cognitive load. Our findings revealed that the presence of engagement cues significantly improved monitoring accuracy, perceived quantity, and perceived teaching behavior adjustment. Additionally, we observed a significant interaction effect between engagement cues and student engagement level on cognitive load. Overall, our research sheds light on the fundamental impacts of student engagement cues for instructors in online teaching environments. Below, we elaborate on our findings and discuss their implications for the development of future student engagement monitoring systems.

Our findings revealed that the presence of student engagement cues significantly enhanced participants' ability to monitor students' engagement with greater accuracy, as demonstrated by reduced monitoring error and heightened perceived monitoring quality compared to when engagement cues were absent. This enhancement was facilitated by the incorporation of individual- and class-level engagement cues in a simplified form, such as using colored visualizations to represent student engagement levels and class engagement rates in a bar graph. Our focus on the core advantages of engagement monitoring interfaces, amidst the variety of metrics and visualization methods existing in prior research (e.g., [2, 12, 17, 18, 28, 30, 34]), emphasizes the essential benefits of incorporating student engagement cues into online teaching systems.

In terms of monitoring quantity, our findings suggest an intriguing pattern where instructors perceived an increase in monitoring activity, yet the actual time spent on monitoring (indicated by the proportion of eye gaze on the student window) significantly decreased when engagement cues were present compared to when they were absent. This finding underscores the efficiency of interfaces that display engagement cues, as they may allow instructors to quickly evaluate student engagement levels without the need for extended observation. This efficiency likely stems from the ability of engagement cues to provide at-a-glance insights into student states, streamlining the monitoring process.

As a result, the improved accuracy and efficiency in monitoring facilitated by engagement cues might have led to perceived enhancements in the ability of instructors to adjust their teaching behaviors effectively. While this study did not directly measure the types and frequency of these adjustments or their efficiency, participants reported feeling more engaged in making adjustments in response to the cues, indicating that engagement cues might enable more timely and appropriate teaching adjustments, especially for students displaying low engagement. Post-task interviews further highlighted that these monitoring improvements not only aided in refining teaching methods but also in strengthening the connection between instructors and their students. This points to a wider application of student engagement cues, suggesting they might serve not just for monitoring behaviors but also as a means to foster instructor-student relationships and enhance the social presence in online learning environments, aspects that are often lacking in online classes [15].

Although engagement cue interfaces generally improved monitoring quality and quantity, their overall effect on instructors' cognitive load was mixed and not uniformly significant. In the post-task interviews, some reported experiencing a decrease in cognitive load, attributing this to being freed from interpreting students' non-verbal cues. In contrast, others felt an increase in cognitive load, burdened by the continuous need to monitor and adapt to fluctuations in student engagement. Notably, our quantitative research revealed an interaction effect on cognitive load, particularly pronounced when engagement cues signaled low student engagement. This scenario was particularly challenging for less experienced instructors who found low engagement signals to be stressful, potentially leading to heightened psychological stress from the pressure to modify their teaching strategies, leading to an increased cognitive load.

These findings underscore the need for a careful consideration of the context and characteristics of engagement cues. In particular, while negative indicators such as low class-level engagement can offer valuable insights, they may also overwhelm instructors. This aligns with Ma and colleagues' suggestions to reduce negative emotional triggers for instructors [17], emphasizing the complex effects of engagement cues on cognitive load or stress.

Feedback from participants on the future design of student monitoring interfaces highlighted the critical need for systems that are customizable to cater to the varying needs, experience levels, and specific contexts of different classes. This customization could include flexible visualization options, such as toggles for engagement indicators and selective focus on certain categories of student engagement. Additionally, there was a notable interest in utilizing this data not just for real-time monitoring but also for conducting retrospective analyses of classes and teaching methods. Overall, the findings suggest a promising research avenue focused on balancing information utility with cognitive load so instructors can customize monitoring systems to fit their individual teaching situations by optimizing engagement cues. Such flexibility would be key for instructors to effectively navigate the complexities associated with adopting engagement cues, ensuring that monitoring systems enhance their teaching experiences without contributing to cognitive load or stress.

6 LIMITATIONS AND FUTURE WORK

Our study, while providing valuable insights, comes with several limitations that highlight areas for future research. The primary limitation stems from the lack of real-time interaction between instructors and students, as we utilized pre-recorded videos to represent students. This design choice, necessary for controlling experimental conditions to isolate factors influencing instructors' monitoring behaviors effectively, inadvertently limited the scope of interactions possible, such as responding to student questions or offering tailored guidance. As a result, the dynamic interplay of immediate student responses and instructor interventions was not captured. To address this, future studies should incorporate real-time monitoring systems within actual classroom settings. This would provide a richer understanding of how student monitoring systems are integrated into online educational practice from both instructors' and students' perspectives.

Additionally, our study's design did not capture the complete range of scenarios encountered in actual online classes, particularly regarding student camera use. Although we simulated an ideal scenario where all students had their cameras on, this does not reflect the reality of typical online classrooms, where camera use varies. Future research should account for the diversity of student engagement, including scenarios where cameras are off, to understand the support instructors' need in these more common situations.

Moreover, the use of eye tracking to assess instructors' monitoring behavior may not always distinguish between engagement monitoring and merely looking, highlighting a need for more sophisticated techniques that can precisely capture instructors' behaviors during classes.

Finally, we hope to broaden our sample in future studies to include instructors with a wider range of demographic characteristics since the current participant pool skews toward younger instructors proficient in digital technology. By involving instructors of varied ages, digital literacy levels, and teaching methodologies, we intend to deepen our understanding of how different instructors use student engagement monitoring interfaces.

7 CONCLUSION

This research investigated the impact of real-time student engagement interfaces on instructors' monitoring effectiveness, teaching behavior adjustments, and cognitive load in online learning settings. Conducting a controlled lab experiment with 20 instructors, we evaluated two interfaces: one with engagement cues and the other without, across varying levels of student engagement. Findings revealed that engagement cues significantly improved monitoring accuracy and enabled more active teaching adjustments. Additionally, it underscored how class context, along with instructors' experience and preferences, affected the cognitive load associated with employing such interfaces. Overall, our findings reaffirm the positive impact of student engagement monitoring interfaces on the quality and quantity of monitoring. It also emphasizes the critical need for customizable interfaces that balance informational utility with cognitive load, proposing design implications that would allow instructors to tailor engagement cues to their unique teaching scenarios, thereby improving the monitoring efficiency and teaching quality in online learning contexts.

ACKNOWLEDGMENTS

We sincerely thank our study participants and reviewers for their invaluable contributions to this research. Furthermore, we would like to thank Jinsu Eun for his assistance with setting up the experiment. This work was supported by the Ministry of Education of the Republic of Korea and the National Research Foundation of Korea (NRF-2021S1A5B8096358) and partly supported by the New Faculty Startup Fund from Seoul National University (#200-20230022).

REFERENCES

[1] Zienab Alrefaie, Mohammed Hassanien, and Abdulmonem Al-Hayani. 2020. Monitoring online learning during COVID-19 pandemic; Suggested online learning portfolio (COVID-19 OLP). *MedEdPublish* 9, 110 (2020), 110.

[2] Sinem Aslan, Nese Alyuz, Cagri Tanriover, Sinem E Mete, Eda Okur, Sidney K D'Mello, and Asli Arslan Esme. 2019. Investigating the impact of a real-time, multimodal student engagement analytics technology in authentic classrooms. In *Proceedings of the 2019 chi conference on human factors in computing systems*. 1–12.

[3] Jered Borup, Charles R Graham, and Jeffery S Drysdale. 2014. The nature of teacher engagement at an online high school. *British Journal of Educational Technology* 45, 5 (2014), 793–806.

[4] Tinggui Chen, Lijuan Peng, Xiaohua Yin, Jingtao Rong, Jianjun Yang, and Guodong Cong. 2020. Analysis of user satisfaction with online education platforms in China during the COVID-19 pandemic. In *Healthcare*, Vol. 8. MDPI, 200.

[5] Isha DeCoito and Mohammed Estaiteyeh. 2022. Transitioning to online teaching during the COVID-19 pandemic: An exploration of STEM teachers' views, successes, and challenges. *Journal of Science Education and Technology* 31, 3 (2022), 340–356.

[6] Raphael A Dourado, Rodrigo Lins Rodrigues, Nivan Ferreira, Rafael Ferreira Mello, Alex Sandro Gomes, and Katrien Verbert. 2021. A teacher-facing learning analytics dashboard for process-oriented feedback in online learning. In *LAK21: 11th International Learning Analytics and Knowledge Conference*. 482–489.

[7] Fiona Draxler, Linda Hirsch, Jingyi Li, Carl Oechsner, Sarah Theres Völkel, and Andreas Butz. 2022. Flexibility and Social Disconnectedness: Assessing University Students' Well-Being Using an Experience Sampling Chatbot and Surveys Over Two Years of COVID-19. In *Designing Interactive Systems Conference*. 217–231.

[8] Jill M Galusha. 1998. Barriers to learning in distance education. (1998).

[9] David García-Solórzano, Germán Cobo, Eugènia Santamaría, Jose Antonio Morán, Carlos Monzo, and Javier Melenchón. 2012. Educational monitoring tool based on faceted browsing and data portraits. In *Proceedings of the 2nd International Conference on Learning Analytics and Knowledge*. 170–178.

[10] Sandra G Hart. 2006. NASA-task load index (NASA-TLX); 20 years later. In *Proceedings of the human factors and ergonomics society annual meeting*, Vol. 50. Sage publications Sage CA: Los Angeles, CA, 904–908.

[11] Mariam Hassib, Daniel Buschek, Paweł W Wozniak, and Florian Alt. 2017. HeartChat: Heart rate augmented mobile chat to support empathy and awareness. In *Proceedings of the 2017 CHI conference on human factors in computing systems*. 2239–2251.

[12] 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.

[13] Denis Helic, Hermann Maurer, and Nick Scherbakov. 2000. Web based training: What do we expect from the system. In *Proceedings of ICCE*. 1689–1694.

[14] Kathleen A Jeffery and Christopher F Bauer. 2020. Students' responses to emergency remote online teaching reveal critical factors for all teaching. *Journal of Chemical Education* 97, 9 (2020), 2472–2485.

[15] Wonjin Lee, Kayoung Park, Jiyoung Yoo, and Hajin Lim. 2023. Design to Improve Social Presence among University Students in Asynchronous Online Learning Environment. In *Extended Abstracts of the 2023 CHI Conference on Human Factors in Computing Systems*. 1–7.

[16] David John Lemay, Paul Bazelais, and Tenzin Doleck. 2021. Transition to online learning during the COVID-19 pandemic. *Computers in Human Behavior Reports* 4 (2021), 100130. https://doi.org/10.1016/j.chbr.2021.100130

[17] 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 *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems*. 1–25.

[18] Kiyosu Maeda, Riku Arakawa, and Jun Rekimoto. 2022. CalmResponses: Displaying Collective Audience Reactions in Remote Communication. In *ACM International Conference on Interactive Media Experiences*. 193–208.

[19] Roberto Martinez Maldonado, Judy Kay, Kalina Yacef, and Beat Schwendimann. 2012. An interactive teacher's dashboard for monitoring groups in a multi-tabletop learning environment. In *Intelligent Tutoring Systems: 11th International Conference, ITS 2012, Chania, Crete, Greece, June 14-18, 2012. Proceedings 11*. Springer, 482–492.

[20] Riccardo Mazza and Vania Dimitrova. 2007. CourseVis: A graphical student monitoring tool for supporting instructors in web-based distance courses. *International Journal of Human-Computer Studies* 65, 2 (2007), 125–139.

[21] Miroslav Minović and Miloš Milovanović. 2013. Real-time learning analytics in educational games. In *Proceedings of the first international conference on technological ecosystem for enhancing multiculturality*. 245–251.

[22] 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.

[23] Anna Ya Ni. 2013. Comparing the effectiveness of classroom and online learning: Teaching research methods. *Journal of public affairs education* 19, 2 (2013), 199–215.

[24] Krystle Phirangee, Carrie Demmans Epp, and Jim Hewitt. 2016. Exploring the Relationships between Facilitation Methods, Students' Sense of Community, and Their Online Behaviors. *Online Learning* 20, 2 (2016), 134–154.

[25] Lawrence C Ragan. 1999. Good teaching is good teaching: An emerging set of guiding principles and practices for the design and development of distance education. *Cause Effect* 22 (1999), 20–24.

[26] Chrysi Rapanta, Luca Botturi, Peter Goodyear, Lourdes Guàrdia, and Marguerite Koole. 2020. Online university teaching during and after the Covid-19 crisis: Refocusing teacher presence and learning activity. *Postdigital science and education* 2 (2020), 923–945.

[27] Maria Rosala. 2019. How to analyze qualitative data from UX research: Thematic analysis. *NN-Nielsen Norman Group* (2019).

[28] Arissa J Sato, Zefan Sramek, and Koji Yatani. 2023. Groupnamics: Designing an Interface for Overviewing and Managing Parallel Group Discussions in an Online Classroom. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*. 1–18.

[29] Dick Schoech and David Helton. 2002. Qualitative and quantitative analysis of a course taught via classroom and internet chatroom. *Qualitative Social Work* 1, 1 (2002), 111–124.

[30] 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.

[31] Anouschka Van Leeuwen, Jeroen Janssen, Gijsbert Erkens, and Mieke Brekelmans. 2014. Supporting teachers in guiding collaborating students: Effects of learning analytics in CSCL. *Computers & Education* 79 (2014), 28–39.

[32] Camilo Vieira, Paul Parsons, and Vetria Byrd. 2018. Visual learning analytics of educational data: A systematic literature review and research agenda. *Computers & Education* 122 (2018), 119–135.

[33] Emma R Wester, Lisa L Walsh, Sandra Arango-Caro, and Kristine L Callis-Duehl. 2021. Student engagement declines in STEM undergraduates during COVID-19–driven remote learning. *Journal of microbiology & biology education* 22, 1 (2021), ev22i1–2385.

[34] Yingying Wu, Yiqun Liu, Yen-Hsi Richard Tsai, and Shing-Tung Yau. 2019. Investigating the role of eye movements and physiological signals in search satisfaction prediction using geometric analysis. *Journal of the Association for Information Science and Technology* 70, 9 (2019), 981–999.