MirrorUs: Mirroring Peers' Affective Cues to Promote Learner's Meta-Cognition in Video-based Learning

SI CHEN, Information Sciences, University of Illinois at Urbana–Champaign, USA
JASON SITU, Computer Science, University of Illinois at Urbana-Champaign, USA
HAOCONG CHENG, Information Sciences, University of Illinois at Urbana-Champaign, USA
DESIRÉE KIRST, Gallaudet University, USA
YUN HUANG, Information Sciences, University of Illinois at Urbana-Champaign, USA

Learners' awareness of their own affective states (emotions) can improve their meta-cognition, which is a critical skill of being aware of and controlling one's cognitive, motivational, and affect, and adjusting their learning strategies and behaviors accordingly. To investigate the effect of peers' affects on learners' meta-cognition, we proposed two types of cues that aggregated peers' affects that were recognized via facial expression recognition: *Locative* cues (displaying the spikes of peers' emotions along a video timeline) and *Temporal* cues (showing the positivities of peers' emotions at different segments of a video). We conducted a between-subject experiment with 42 college students through the use of think-aloud protocols, interviews, and surveys. Our results showed that the two types of cues improved participants' meta-cognition differently. For example, interacting with the *Temporal* cues triggered the participants to compare their own affective responses with their peers and reflect more on why and how they had different emotions with the same video content. While the participants perceived the benefits of using AI-generated peers' cues to improve their awareness of their own learning affects, they also sought more explanations from their peers to understand the AI-generated results. Our findings not only provide novel design implications for promoting learners' meta-cognition with privacy-preserved social cues of peers' learning affects, but also suggest an expanded design framework for Explainable AI (XAI).

.

CCS Concepts: • **Human-centered computing** → **Empirical studies in HCI**; **Computer supported cooperative work**; • **Applied computing** → **Distance learning**.

Additional Key Words and Phrases: Online Learning, Facial Recognition, Meta-Cognition

ACM Reference Format:
Si Chen, Jason Situ, Haocong Cheng, Desirée Kirst, and Yun Huang. 2023. *MirrorUs*: Mirroring Peers' Affective Cues to Promote Learner's Meta-Cognition in Video-based Learning. *Proc. ACM Hum.-Comput. Interact.* 7, CSCW2, Article 288 (October 2023), 25 pages. https://doi.org/10.1145/3610079

Authors' addresses: Si Chen, sic3@illiois.edu, Information Sciences, University of Illinois at Urbana–Champaign, Champaign, Illinois, USA, 61802; Jason Situ, junsitu2@illinois.edu, Computer Science, University of Illinois at Urbana-Champaign, Urbana, Illinois, USA; Haocong Cheng, haocong2@illinois.edu, Information Sciences, University of Illinois at Urbana-Champaign, Champaign, Illinois, USA; Desirée Kirst, deskirst@gmail.com, Gallaudet University, Washington, District of Columbia, USA; Yun Huang, yunhuang@illinois.edu, Information Sciences, University of Illinois at Urbana-Champaign, Champaign, Illinois, USA.

2573-0142/2023/10-ART288 $15.00
https://doi.org/10.1145/3610079

1 INTRODUCTION

Meta-cognition refers to the awareness and control of one's cognitive, motivational, and emotional states (also called affects), and adjusting their own learning strategies and behaviors [34, 49, 68, 69]. Recent studies have begun to explore meta-cognition in social and online learning environments, and have identified important meta-cognition processes such as sharing ideas and inviting peers to evaluate them [69]. To facilitate meta-cognition for improved learning experiences and reflections, tools are needed to raise learners' awareness about their own and peers' learning affects [13, 49]. For example, Automatic Emotion Recognition (AER) via facial movements is an Artificial Intelligence (AI) technology and can be applied to detect learner's affective states at low expense [3, 15, 48].

However, there is a lack of research on sharing AER-based affects and little is known about the effect of using AI-generated peers' affects for improving individuals' learning experiences. Also, with regard to AI-based solutions, privacy, and trustworthiness are important ethical factors to consider [15, 41, 51]. It is crucial to understand learners' willingness to share their AER-based affects and how the potential ethical concerns could impact their perceptions and use of the AI-generated affects of themselves and their peers.

To address the voids, we designed and prototyped *MirrorUs*, an AER-based reflection tool for reflecting on video-based learning. It presents learners' own affects with peers' affects that are generated by an AER-based technology via facial movements. We named the tool *MirrorUs*, hoping that learners can be more aware of their own learning affects by mirroring others'. Through a need-finding study and drawing from existing literature, we proposed two types of AI-generated affect cues that aggregated peers' affects: *Locative* cues that display spikes in peers' emotion along the video timeline and *Temporal* cues that show the positivities of peers' emotions at different segments of a video. We conducted a between-subject study with three groups of participants: *Locative* (N=15), *Temporal* (N=14), and the control group (N=13) to evaluate the two designs. We examined participants' think-aloud, interaction with the affects cues, interview feedback, and survey input.

Our work makes significant contributions to the CSCW community. First, we provide new empirical evidence showing that AI-generated peers' affects could trigger three types of meta-cognition processes in reflection, e.g., *(dis)agreement with peers affects*, *recalling content to explain own affects*, and *building shared knowledge*. Specifically, *Temporal* cues had a positive effect on improving (social) meta-cognitive processes. Second, our findings provided new insights into collecting human input to explain AI-generated results. Specifically, participants looked for human input for explaining the AI-generated affects so as to improve participants' trust in the tool. This finding suggested an expanded explainable AI design framework proposed in [25] by adding a "How" category (e.g., explaining how the AI results are generated) to the existing 4W (What, why, who, and when) categories for improved social transparency of AI-based systems.

2 RELATED WORK

2.1 Learner's Affects and Meta-Cognition in Online Learning

Meta-cognition allows learners to become more aware of their own learning processes and to adjust their own learning strategies and behaviors [46, 49]. Research has found that by improving learners' awareness of their own and peers' affects, such as engagement, flow, curiosity, frustration, and boredom [11], meta-cognition can be increased. This awareness can also help learners understand and monitor their own learning performance [15, 21, 23, 24]. Additionally, observing how peers respond to a certain situation can create a feeling of curiosity that encourages learners to think more deeply about how their actions differ from those of their peers [24], which can be beneficial

for meta-cognition. Our goal is to increase meta-cognition by improving awareness of one's own and peers' affects.

A growing body of research is investigating the impact of meta-cognition in online learning environments. Studies have demonstrated that meta-cognitive strategies, such as self-regulation, can lead to greater academic success and improved learning outcomes in offline settings [19]. Research has further argued that meta-cognition is beneficial for enhancing social learning in online environments, with self-regulatory strategies promoting collaborative learning success [22, 66]. Chan [13] investigated the social aspects of meta-cognition and demonstrated that computer-supported collaborative learning environments help learners conduct social meta-cognition to be aware of knowledge differences between themselves and other group members. However, few studies have investigated the impact of promoting meta-cognition with peers' affects in online learning settings, particularly with Automatic Emotion Recognition (AER) technology (explained in the following section).

2.2 Automatic Emotion Recognition (AER) in Online Learning

With the advance of AI, Automatic Emotion Recognition (AER) technologies have been widely used in both research and industry settings. AER is a sub-field of AI that involves the development of algorithms and models that can automatically detect and interpret emotions. AER technologies detect emotion using input channels such as body movements [65], speech [59], and brain activities [1]. Recent AER technologies build on facial recognition models to predict affective states based on images [58, 60]. AER technology for analyzing and visualizing facial movement and detecting learners' affects is an important part of online learning environments. For example, AER via facial movements helps teachers become more aware of learners' emotions and behaviors in online learning [26, 44, 47], to conduct online communication skills training [50, 56], and to proctor online exams [41]. More importantly, visualizing emotions identified by AER via facial movement in an online learning system can raise learners' self-awareness and support reflections [3, 14, 15, 48].

Learners' affects can be represented in categorical emotions (e.g. surprise, enjoyment [33, 39]) or dimensional spaces (e.g. arousal and valence features [6, 57]). Russell et al. [54] introduced the circumflex model which categorizes emotions into two-dimensional arousal-valence circle. An arousal scale ranges from -1 (extremely calm) to 1 (extremely excited) that measures the intensity of excitement, while a valence scale ranges from -1 (extremely negative) to 1 (extremely positive) that measures pleasantness [6]. AER technology is able to detect learners' emotional states through facial movements in both categorical and dimensional affects in online learning. Dimensional affects are preferred for dynamic analyzing emotional states' intensity and transitional behaviors. To illustrate, the self-regulated learning systems *MetaTutor* and *Mirror* used dimensional space to track learners' performance and awareness of their own affects, respectively. Chen et al. [15] found that when learners used a two-dimensional model rather than a categorical model for self-reflection, they had more freedom to interpret their own emotions due to the imperfection of AER data. In this work, we utilized a facial recognition model (MTCNN API [20]) to represent learning's affects in dimensional spaces (arousal and valence feature).

2.3 Explainable and Ethical AER-based Online Learning Systems

The primary goal of Explainable AI (XAI) aims to provide explanations to make AI models' actions and decisions understandable to humans [43]. In the context of learning systems, most of the existing studies in XAI are model-centered rather than human-centered [25], namely, they do not address the emotional responses of humans as they interact with the explanations [29]. Ehsan et al. [25] proposed a human-centered perspective that takes into account sociotechnical factors and

considers the 4W contexts (What, Why, Who, and When) to improve the social transparency of AI systems by explaining how they mediate decision-making.

In intelligent learning systems, some recent works started to explore learner-centered approaches to explain AI-based learning systems. For example, Afzaal et al. employ techniques in learning analytics with XAI models to provide intelligent feedback and aid in promoting self-regulation of learning [2]. Conati et al. designed an explanation function with personalized hints in the intelligent tutoring system and found that their design helped students use learning tools in open-ended learning environments, by making computational models more transparent and interpretable to learners [18]. However, these explainable AI-based learning systems have not yet integrated AER technologies nor have they been implemented for online learning systems.

In order to effectively implement AER-based online learning systems, it is crucial for AI to provide explanations that are understandable to both educators and learners. Madumal et al. mentioned that XAI or explanations has the potential to resolve ethical issues related to the actions, decisions, and behaviors of AI in online learning systems towards learners [45]. For instance, ethical concerns include but are not limited to, the accuracy of the AER technology along with privacy concerns [15, 31, 51]. Learners have expressed concerns about not being able to control whether the system is detecting their emotions and if their data is shared with others [15, 47]. Sharing emotion-sensitive information could raise privacy concerns and transparency issues [37, 41, 67]. Moreover, the access to data collected by the emotional system raised risk of surveillance on both students and tutors [9, 41, 51]. From the technical perspective, there have been concerns about the accuracy of facial expression recognition algorithms, namely, users' own interpretations of the emotions may not be accurately perceived [4, 28, 53] and how these systems handle and distribute users' images obtained through facial expression recognition algorithms [47, 56]. In summary, limited studies are made on understanding and addressing ethical concerns of XAI for these online learning systems [36].

2.4 Research Questions

In Section 2.2, we discussed how AER technology has exhibited the potential in identifying learners' emotional states. Our first research question (RQ1) aims to explore how these recognized affects can aid learners' metacognition in online learning when utilizing peers' affects. However, it's crucial to note that AER technology requires transparency and raises ethical concerns, as stated in Section 2.3, which includes disclosing data, algorithms, and resulting outputs. Therefore, our second research question (RQ2) focuses on investigating the impact of ethical concerns on learners' willingness to share their detected affective states with others. This study aims to analyze the interaction between learners and their peers' affective states in RQ1 and to understand whether learners are willing to participate in the ecosystem by sharing their affective states with their peers in RQ2.

- **RQ1**: *How do learners reflect on their learning experience by navigating peers' affects in MirrorUs?*
- **RQ2**: *What are learners' perceptions of sharing own affects in MirrorUs for peers' reflection?*

3 *MIRRORUS* SYSTEM DESIGN

To answer our research questions regarding the effectiveness of reflecting on peers' AER-based affects and perceptions of sharing their own AER-based affects for other system users, we conducted two need-finding sessions. In *Round I*, our goal was to explore what types of peer affects learners would find useful in video-based learning. In *Round II*, our goal was to identify the system features that supported the representation of peers' affects and were preferred by learners for interaction and reflections. We proposed two types of peers' affects leveraging AER via facial movements, *Locative* cues (displaying peers' *emotion spike* along the video timeline) and *Temporal* cues (showing

the positivities of peers' emotions at different segments of a video) shown in Figure. 1. These two types of peers' affects were evaluated in RQ1 and RQ2. Our system supports video-based learning and the whole system flow includes individual video watching, self-reporting learning moments, and reflection with own and peers' affects as shown in Figure 2.

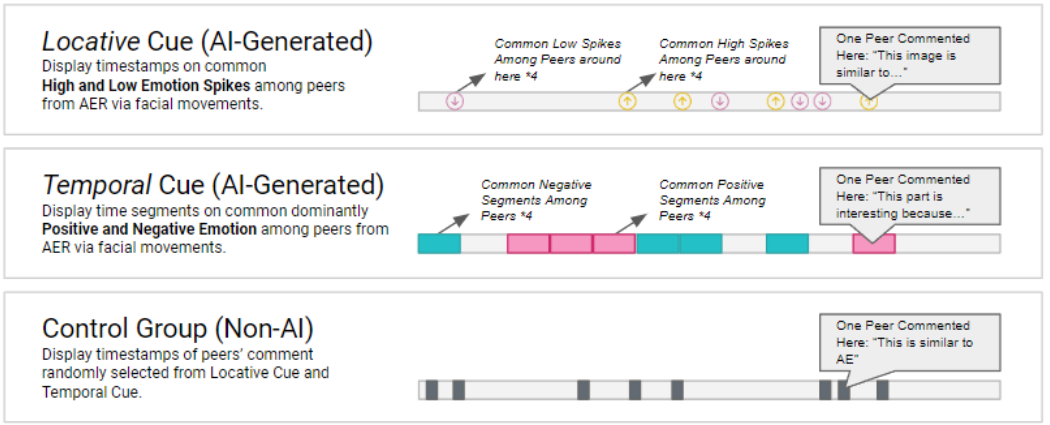

Fig. 1. Different Peers' Cues Presented in *MirrorUs* for Reflections among Three Groups. One treatment group displayed AI-generated *Locative* cues: aggregated peers' *emotion spikes* from AER; one treatment group presents AI-generated *Temporal* cues: aggregated peers' *dominant emotions* from AER as shown in Figure 1. The control group only displayed peers' comments without peers' affects. The feature of displaying peers' comments to facilitate reflection is the same across all three groups informed by need-finding studies. In our evaluation study (RQ1 and RQ2), the peers' comments and the AER were mock-up data, as explained in Section 3.2.2.

3.1 Need-finding in Two Rounds

3.1.1 Round I - Identifying Different Types of Peers' Affects Useful for Reflections. For this activity, participants were asked to describe to the researcher what types of peers' affects they were interested in interacting with for reflection purposes by viewing their own and one peer's emotion intensity graphs generated from AER via facial movements. 16 participants were recruited for this activity (10 female, 6 male, ages 19-28, students in higher education, F1-F16).

Study Process: Participants were first asked to watch a 15-minute video on the topic of Augmented Reality. In the next step, participants were asked to provide text-based comments at self-selected video timestamps regarding their learning moments and affects, also denoted as "self-report" in the rest of the paper. Participants were then asked to reflect on their learning experience through the think-aloud protocol while engaging interactively with a graph showing the emotional intensity of a peer and themselves (as shown in Appendix A.2). The peer learner's graph had been collected by a researcher prior to the activity and was presented to all participants. Think-aloud is commonly used by researchers to collect data on how users understand visualization. e.g. [17, 55]).

System Details of Generating Own and Peer's Emotion Intensity Graph from AER via Facial Movements: We created an emotional intensity graph by recording facial images of a learner during the viewing of an online video and applying the state-of-the-art regression algorithm (MTCNN [20]) to predict the arousal-valence pairs. The system stores the captured images on the learner's computer and then sends them to the remote server for AER via facial movements. To map the arousal-valence

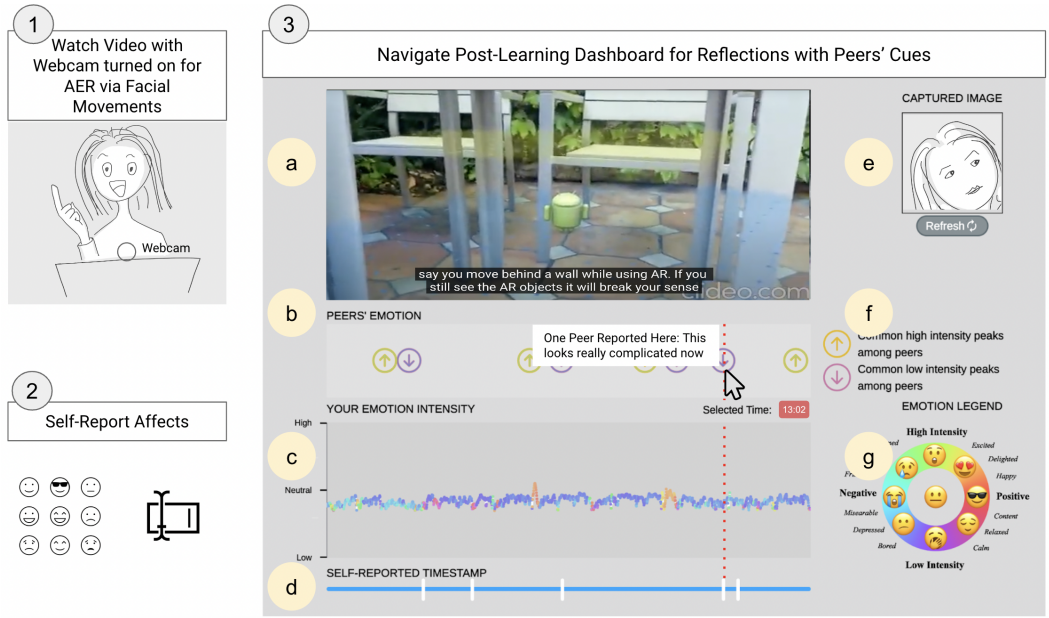

Fig. 2. The User Interaction Flow of *MirrorUs*: (Step 1) The system uses a webcam to detect facial movements and interpret user affect from AER while the user is watching the video; (Step 2) user self-reports their affects **after** watching the video through emoji and text; (Step 3) user reflects with their own effects with additional peers' affects on the post-learning dashboard. Post-learning dashboard has seven areas: (a) displaying the video recording, (b) displaying peers' cues and **differ among the three groups** (in our study, text comments are mocked-up with previous learners' self-reports) (c) visualizing own time-based AER data via facial movements, (d) tagging self-reported learning moments, (e) providing video selfies captures at the selected timestamp (illustrated according to a captured image by researchers to preserve participants' privacy; video selfies did not automatically show up and users needed to click 'Refresh' for it to show up (f) providing a legend for time-based AER data of own affects, and (g) providing a legend for peers' affects (this is not shown for the control group). The sample data was from L14.

pairs into visual graphics, we generated a 2D scatter plot by using the HSL color wheel to color-code each point according to its angle from the arousal-valence 2D values. Additionally, we provided a corresponding legend of arousal-valence 2D space on the interface to interpret the affects. The legend uses a continuous rainbow with cool colors representing negative results and warm colors representing positive results [15, 32, 40] to depict the relationship between arousal and valence. The emotional intensity graph and its legend were implemented the same in the formal study as shown in Figure 2 area (c) and (g). To ensure the privacy of the data, the process of predicting emotional data from images was handled on the remote server, while the process of capturing own facial images is stored locally.

All 16 participants reflected on their interactions on their own and peers' emotion intensity graphs. A total of 47 interactions were recorded in relation to the video content reflection. In particular, interest in mentioning peers' positive/negative segments was reflected about 28 times, while 19 reflections emerged regarding peers' arousal. Our analysis of participants' feedback revealed two cues that facilitate time-wise comparison: mostly positive/negative, and observed emotion spike. Participants intended to explain why a video segment had dominant positive or

negative emotions as well as compare whether their peers had similar experiences. Our think-aloud transcripts illustrated how participant F11 tried to find common affects from the peer graph. *"In the beginning, there was a change from orange to yellow, and I know it was like a mix between curiosity, delightful and recognizing fresh research. Seems like this peer felt similarly **mostly positive**, as I did."*

Participants were also interested in reviewing video content that caused spikes in the emotion intensity graph. Participants wanted to determine whether they had overlooked any content that was particularly intriguing or confusing as the spikes were perceived as being caused by interesting content. For example, F9's think-aloud revealed the motivation to rewatch a specific segment in the video. *"Oh there is a **peak** around the middle part, but I didn't have any spikes here. Maybe I missed something here [Rewind Video]. I see, I also found it very exciting and really attracted my attention. I actually wanted to take a self-report for this part while watching the video but I forget, it was about the AR for NGO."*

Findings: Two Types of Affects via AER for Post-Learning Reflections Recent works on AER have proposed two metrics to capture transient affects: *emotion spikes*, which are brief and intense changes from the expected mean emotion; and *dominant emotions*, which are the most prevalent emotion labels throughout a given time-frame [35]. Our findings from *Round I* align with these metrics. Participants observed sharp changes in the emotion intensity graph, which correspond to *emotion spikes*, and the use of mostly warm/cool colors to represent positive/negative emotions in a time segment, which correspond to *dominant emotions*. Consequently, we plan to propose two types of peers' affects reflecting *emotion spikes* and *dominant emotions* based on AER.

Findings: Self-Reporting After Watching Videos for a Focused Learning Experience: Participants suggested that self-reporting and watching the video should be done separately. Self-reporting while watching a video would require pausing and resuming, particularly when the video is 15 minutes long. This process involves recalling and articulating one's feelings about the video content, which might disrupt the learning process. Therefore, our workflow consists of three steps: watching a video, providing self-reported learning moments, and navigating affective cues for reflection. This is illustrated in Figure 2.

3.1.2 Round II - Features to Support Reflections using Peers' Affects. This study was designed to understand features that can support reflection on the peers' affects. The same 16 participants of *Round Study I* were invited to this study one month later, and ten of them participated.

Study Process: First, participants marked out positive/negative *dominant emotions* and timestamps with *emotion spikes* on their own emotion intensity graph from *Round Study I*. This was done to help them become familiar with their own affects data, and for researchers to collect participant affects to aggregate peers' affects cues in the formal study. Secondly, participants viewed four prototypes to stimulate their thoughts and consideration when comparing their own and peers' affects. We focused on designs that would help with effectively comparing 2D arousal-valence affects, as prior research suggested that users find it challenging to interpret these data together in the 2D scatter space [15, 27]. An example of prototype: an interface highlighting sections across two time-series line graphs of own and peers' affects, which is the AI-recognized differences.

Feature-1: Aggregated Affects for Anonymous Comparison: Participants expressed strong concerns about privacy when it comes to sharing affective data at a small granularity, such as their own affects. They suggested that even anonymous own affect graphs should not be shared. They also considered personal emotion eliciting privacy issues and not as valuable for other learners. Instead, they felt that the most beneficial data for other learners is the collective affects of the whole class, which should be shared in an aggregated form.

Feature-2: Parallelize Own Affects and Peers' Cues for Time-wise Comparison: Wile participants acknowledged the potential of system-generated suggestions for comparing different expressions, they expressed skepticism as to the accuracy of using the system to define AER differences. To ensure accuracy, participants preferred to manually compare AER in the context of their own emotional expressions. This study focused on giving users the ability to make time-wise comparisons themselves, rather than exploring designs that would increase the perceived accuracy of AI.

Feature-3: Prefer Reading Text Comments to Facilitate Recall. Participants indicated that providing textual comments as a means of interpreting and verifying the affects derived by the AER models would be helpful. Some participants noted that seeing textual comments can assist them in recalling video content they might have overlooked. Participants also realized that text comments can help them compare their own interpretations of video content with others'. In other words, participants wanted to see text comments from their peers to identify whether their learning experiences were similar or discrepancies.

3.1.3 *Proposing Two Types of Cues Aggregating Peers' Affects based on AER via Facial Movements.*

Here, we propose **Locative** cues and **Temporal** cues as a result of taking account into *Round 1* takeaway on different types of affects participants perceived useful, and *Round II* takeaway that aggregating the peers' affects (*Feature-1*) along the timeline (*Feature-3*) and providing text comments (*Feature-3*), as illustrated in Figure 1. The peers' cues in the control group had no information about peers' affects and are only text comments to facilitate recall (*Feature-3*) along the video timeline (*Feature-2*).

We named the aggregated peers' common *emotion spikes* as *Locative* cues, while the peers' common *dominant emotions* were named as *Temporal* cues. In the *Locative* cues, *emotion spikes* are displayed with up and down arrows to represent peers' high and low arousal. In the *Temporal* cues, *dominant emotions* are presented using segments in the video timeline, with positive and negative emotions highlighted with colors Pastel Magenta and Teal respectively. We integrated the parallelize own affects and peers' cues for time-wise comparison (*Feature-2*) by arranging the peers' cues vertically above the own affects (as shown in Figure 2).

3.2 Finalized *MirrorUs* Design to Support Reflection using Affects

Our major workflow consists of three steps: watching a video, providing self-reports, and navigating affective cues for reflection. Such learning processes using the *MirrorUs* system are illustrated in Figure 2. Step 1, step 2 and step 3 in Figure 2 are detailed in Section 3.2.1. In Section 3.2.2, we described how we mocked-up affects data to support consistency across participants in terms of *Locative* and *Temporal* cues. Our mock-up approach to simulating peers' affects is not the only possible approach. Rather, it is one of the methods that emerged through our two rounds of need-finding.

3.2.1 *MirrorUs Workflow.*

Step 1: Collecting Own AER data while Watching Video: In our system, a state-of-the-art regression model (MTCNN [20]) is used to detect individual affects using facial images. The system captures continuous image streams, locates human faces, crops the images into square shapes that contain only the learner's head region, and then feeds the resulting images into the model for predicting arousal-valence values. The arousal-valence values are visualized as emotion as an intensity graph to appear in Figure 2 area (c).

Step 2: Self-Report Affects: After watching the video, participants were presented with a table containing a timeline of the video and were asked to write a reflection associated with each point in the timeline. They were also asked to select an emoji from a list of nine options that represented *confusion, frustration, surprise, curiosity, delight, flow, fresh research, bored, neutral.* These emoji

options were presented in an arousal-valence matrix, which has been used in previous models to predict affective states in learning [33, 39]. Once the participants provided their comments and selected an emoji, the affective states were pinned onto the timeline shown in Figure 2 area d.

Step 3: Reflection with Own and Peers' Affects After Watching Videos MirrorUs features two variations of time-wise comparison, as illustrated in Figure 1. The first one presents peers' cues of high and low *emotion spikes*. The individualized graphs help learners make sense of their own affects, especially changing between different emotional states. Learners are able to compare their own affects with those of their peers by clicking anywhere on their own graph. A vertical red dotted line is displayed in both own and peers' affects to enable comparison.

Self-report is the same process and feature that exists in all three groups. A Kruskal-Wallis test and Dunn's test analysis indicated that participants' self-report in step 2 quantity in formal evaluation (RQ1 and RQ2) show no statistical significance difference between the three groups (M= 6.9, SD=3.3). Also, in the think-aloud session, the majority of participants did not click self-report (on average 0.10 time per participant) in step 3. Therefore, we do not focus on self-report in the main finding sections and reflect on its design in the discussions.

3.2.2 Mock-Up Data of Locative and Temporal Cues. In this study, we mocked-up peers' affects using the data from ten participants that attended both rounds of need-findings. We introduces our mock-up process below. Kaur et al. presented an approach for identifying *emotion spikes* and *dominant emotions* by using grid searches and algorithms, and its outputs were then confirmed by human participants [35]. Instead of confirming with participants, our research directly used each participant's manual mark out *emotion spikes*, *dominant emotions* collected in *Round II* of the need-finding and researchers aggregated the affects data.

In order to display high-realistic peers' affects, researchers first reviewed all intensity graphs that included text labels of emotions on multiple video segments that participants in *Round II* had manually labeled. For each 1-minute interval, the percentage of participants that marked it as a positive segment, negative segment, high intensity peak and low intensity peak was added up. The top eight segments with the highest percentage of more than 50% participants expressing positive or negative were marked as "common agreed positive/negative" and displayed as *Temporal* cues, while the top eight agreed high/low peaks were chosen to display in *Locative* cues. The reason for selecting the top eight was that in *Round I*, the average self-reports per person was 7.5, thus most participants were comfortable with a cognitive load consisting of reflections across eight timestamps. Then for peers'text comments, The researcher screened 87 self-reported texts to identify the ones that *match* the chosen time intervals and the AER data to present as peers' comments. Eight qualified peers' comments were selected for each time interval in *Locative* cues; Eight qualified peers' comments were selected for each time interval in *Temporal* cues. As for the control group, four from *Locative* cues and four from *Temporal* cues were displayed. The peers' comments were displayed anonymously to ensure comfort for sharing.

4 METHOD

4.1 Study Design for *MirrorUs* Evaluation

To understand how learners would use *MirrorUs* for video-based learning, we conducted a three-session user study, as illustrated in Figure 3. We recruited 42 participants (21 female, 20 male) and randomly assigned them to one of the three groups. The control group has 13 participants, while the group reflecting with *Locative* cues and *Temporal* cues had 15 and 14 participants each. In the remainder of this paper, we denoted the participants who were provided with the *Locative* cues by L01-L15 and the participants with the *Temporal* cues by T01- T14. Participants from the control group were denoted as C01-C13.

All participants are current college students, including 16 who reported pursuing a graduate-level degree. Of the 42 participants, 29 were majoring in information sciences, computer science, or engineering, while the other participants were majoring in education, mathematics, linguistics, statistics, physiology, or food science. As for ethnicity, 29 out of 43 participants described themselves as Asian, while six are White, five are African, one is American Indian, one is Mixed, and one participant preferred not to share ethnicity. Note that all participants in this round of evaluation did not participate in our previous studies discussed in Section 3, nor did they know the prior studies. None of them reported accessibility needs in their everyday learning. We compensated each participant 10 USD per hour. Each study took about two hours (Figure 3).

We used the same 15-minute video on introducing Augmented Reality from the need-finding study and recruited participants with no prior background in it. While they were watching the video, individuals were asked to turn on their webcams with consent granted in advance, and then self-report their feelings. We also asked them to reflect on their learning experience using the post-learning dashboard and participate in an exit survey and interview to answer RQ1 (think-aloud, survey, and interview data) and RQ2 (survey and interview data). The study was approved by our university's Institutional Review Board.

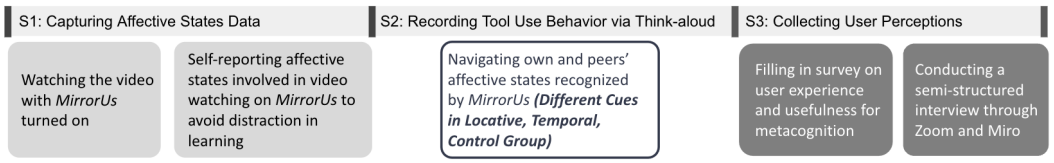

Fig. 3. Three sessions were included in the study: 1) capturing participant's affective states from two methods, i.e., automatically recognized by the *MirrorUs* and self-reported by the participants; and 2) recording use behaviors via think-aloud; and 3) collecting participants' perceptions of the tool-use through survey and interview.

4.1.1 Recording User Behavior: Think-Aloud using Post-Learning Dashboard (Verbalization + Interaction with MirrorUs). The second session followed prior research on users' interactions with personal data visualizations [17]. A researcher first demonstrated the *MirrorUs* interface via screen-share on Zoom. Afterwards, the researcher asked the participants to 'think aloud' as they explored the interface on their own computers. The researcher observed the participants' interactions with the interface and asked them to describe their observations. Each participant's reflection lasted 5-15 minutes. Two participants' think-aloud was not fully recorded due to internet connection issues and was excluded from the analysis. No participants reported difficulties with think-aloud.

4.1.2 Collecting User Perceptions: Survey and Exit Interview. The exit survey was designed to measure participants' perceived usefulness for meta-cognition, cognitive demand, and perceptions of using *MirrorUs* to reflect on their video-based learning experience. All questions used a seven-point Likert response scale (1 *Strongly Disagree* - 7 *Strongly Agree*). The first part of the survey included ten questions on *Perceived Usefulness for Meta-cognition* adapted from existing meta-cognitive scale MCQ-30 [61] and two questions on *Cognitive Demand* using *MirrorUs*. The second part of the survey included six questions on the perception of sharing own affects with peers users using *MirrorUs* informed by ethics of AI in education framework [31]. Trust was measured using two questions, from both the increase and decrease side, inspired by [10] that suggested increased trust and reliance in computing systems is not necessarily beneficial. This session lasted between 25 minutes to 50 minutes. In the exit interview, we asked participants questions on user experience

and concerns of interacting with *MirrorUs*. Note that due to schedule conflicts, four participants did not take the exit interview.

4.2 Data Analysis

We conducted a think-aloud study to understand reflection using visualizations [17]. We first established properties of the data through open coding, followed by axial coding using the framework of meta-cognition and social meta-cognition developed by Chiu & Kuo [16]. Two researchers independently read and coded 50% of the transcripts, and then they discussed and compiled their codes together via three rounds of discussions. The inter-rater agreement after each code-book refinement was 83%, 92%, and 97%. Then one researcher individually coded the rest of the 50% data. We also analyzed the effect of group membership on verbalizing meta-cognition processes and survey questions using statistical models. Finally, we conducted thematic analysis [8] on participants' interview feedback and identified several themes, such as "effectiveness and challenges of reflection with peers' cues," " suggestions to improve effectiveness," "concerns in sharing affects while reflection," and " suggestions to address concerns."

5 FINDINGS

5.1 Peers' Affects: Locative and Temporal Cues Promoting Meta-cognition (RQ1)

To examine the types of meta-cognitive processes that participants conducted using the peers' cues, we annotated their think-aloud data. Fig 4 shows one example of the coded think-aloud data from T03. The results were then triangulated with participants' survey results about their perceived usefulness and cognitive load of using the two different cues (5.1.1). We further qualitatively investigated interview results to understand *why and how* the two proposed cues impacted the participants' meta-cognition (5.1.2).

5.1.1 Three Meta-cognitive Processes Triggered by Peers' Cues. During the think-aloud session (as shown in Figure 3), each participant interacted with *MirrorUs* and reflected on multiple frames/segments of the video. Participants' spoken sentences allowed us to identify three types of meta-cognitive processes below:

- *(Dis)agreeing with Peers* - Participants only encapsulated whether they (dis)agreed with and/or understood the peers' affects without further explanations (① in Figure 1).
- *Recalling Content to Explain Own Affects* - Participants recalled and retrieved video content to evaluate and explain their own affects (② in Figure 1).
- *Building Shared Knowledge* - Participants explained how and why they (dis)agreed with and/or (fail to) understand peers' affects in detail (③ in Figure 1).

As shown in Figure. 5, statistical analysis of think-aloud data revealed *Temporal* participants verbalized more on *building shared knowledge* than participants in the control group, while no significant group differences were found in the two remaining meta-cognitive processes. Specifically, one-way ANOVA testing on the *building shared knowledge* process indicated there is a significant difference for the three groups ($F(2, 35)=23.6$, $p<.001$). Post-hoc comparisons using the Tukey HSD test indicated that the mean score for *Temporal* participants ($M=7.8$, $SD=4.7$) was significantly higher than the control participants ($M=2.7$, $SD=2.9$), with the power analysis resulted in a value equal to 0.8. Post-hoc comparisons found no significant differences between *Locative* participants and control participants ($p=.077$), *Locative* participants and *Temporal* participants. Before and after removing outliers, the results are the same.

The survey results confirmed that the two cues significantly improved participants' perceived usefulness of the tool for supporting their awareness of their own thoughts (i.e., meta-cognition).

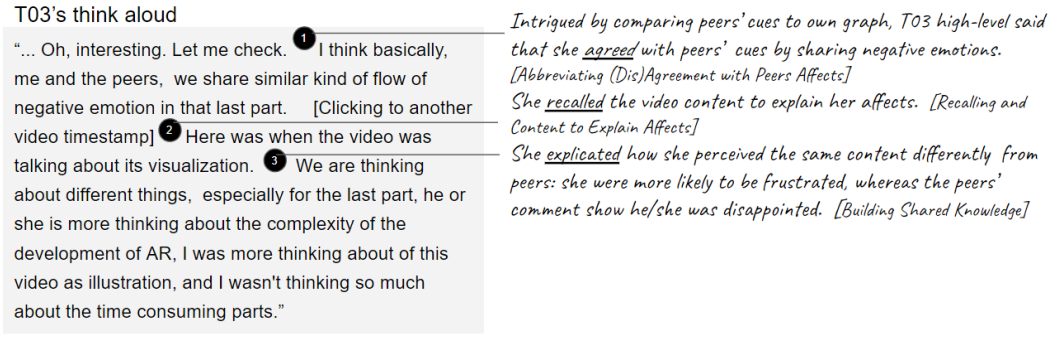

Fig. 4. A sample of think-aloud analysis for meta-cognition: the left image presents a think-aloud transcript from T03; the right annotations show how the researchers coded the transcription in terms of different meta-cognition processes, e.g., *Building Shared Knowledge*.

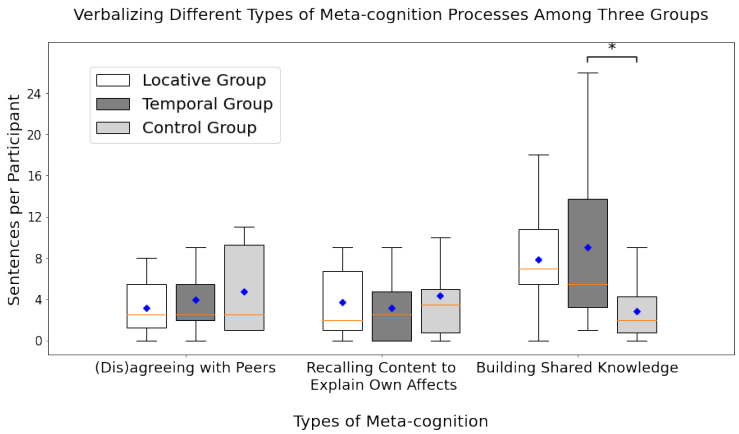

Fig. 5. Boxplot of sentences that participants verbalizing different types of meta-cognition processes among three groups when using *MirrorUs*. Participants in *Temporal* groups verbalized significantly more sentences in *Building Shared Knowledge* than participants in the control group ($p < .05$, denoted *). There are no significant differences among the three groups in *(Dis)agreement with Peers Affects* and *Recalling Content to Explain Own Affects*

Specifically, when rating "This tool can help me constantly examine my thoughts in video-based learning," the *Locative* participants (M=5.7, SD=1.4) and *Temporal* participants (M=5.4, SD=1.1) gave significantly higher scores than the participants of the control group (M=4.2, SD=1.2, $F(2,39) = 0.9$, $p < .01$), with the power analysis resulted in a value equal to 0.92 calculated using existing package [12]. There are no differences between *Locative* and *Temporal* cues.

Participants in the two treatment groups also found it much easier to understand their own affects than the participants in the control group. ANOVA and Post-hoc analysis revealed a significant difference ($F(2,39)=6.306$, $p <.01$) among the three groups, with the power analysis resulted in a value equal to 0.80. *Locative* participants (M=3.1, SD=1.6) ($p<.05$) and *Temporal* participants (M=2.4, SD=1.2) ($p<.01$) found understanding their own affects fairly easy, while those in the control group (M=4.5, SD=1.7) found understanding their own affects moderately difficult. Participants in all

	(Dis)agreeing with Peers	Recalling Content to Explain Own Affects	Building Shared Knowledge
Locative Group			
Peers' Affects	**0.744** **	0.288	0.511
Own Affects	0.476	0.524	0.131
Temporal Group			
Peers' Affects	0.445	**0.894** ***	**0.820** ***
Own Affects	0.526	0.062	0.262

Table 1. Correlations between meta-cognition processes and participants' interactions with AI-generated affects, i.e., peers' and own affects: clicking their own graphs for AI-generated affects (area 3 in Figure 2)); clicking AI-generate peers' cues (area 2 in Figure 2). We observed that *Temporal* and *Locative* had different effects on how participants verbalized meta-cognition processes. Clicking peers' *Temporal* cues encourages participants to recall video content when making comparisons between their own positive and negative affects and those of their peers. This helps them to *build shared knowledge*. The table only displays the data from the two treatment groups, as the control group only showed peers' text comments without peers' affects cues. Significant results with $p < .01$ are denoted **, and results with $p < .001$ are denoted ***.

three groups found understanding peers' cues easy with no significant differences found among the groups.

5.1.2 Two Types of Peers' Affect Cues Leading to Different Meta-cognition Processes. Though both cues triggered participants' three meta-cognition processes, we found that how they support these processes vary. Specifically, we first conducted Pearson's correlation tests to assess the relationship between clicking their own individual graphs for AER affects (area 3 in Figure 2)) (M=2.2, SD=2.5) and the quantity of verbalizing *(dis)agreement with peers*, *recalling content to explain own affects* and *building shared knowledge*. Then we conducted Pearson's correlation tests to assess the relationship between clicking peers' cues to observe how peers perceived the same video (area 2 in Figure 2) (M=15.0, SD=6.5) and the three meta-cognition processes.

The correlation results are shown in Table 1. We observed *Temporal* and *Locative* cues had different effects on how participants verbalized meta-cognition processes, and we interpret them by triangulating think-aloud and interview data qualitatively below. There is no significant correlation between clicking own affects and verbalizing three meta-cognition processes among both two proposed designs.

***Locative* cues associated with meta-cognition of (dis)agreeing with peers.** We found a strong and positive correlation between clicking peers' affects and verbalizing *(dis)agreeing with peers* ($r = .744$, $p < .01$), with the power analysis resulting in a value equal to 0.9. *Locative* participants used peers' affects spikes as hints to visually mark out timestamps' where they have their own emotion spikes, then they were inspired by peers' spikes to find alignment between self and peers. Either their own peaks or peers' spikes intrigued users expressing (dis)agreement with peers; a sample think-aloud quote from L11 is *"I see a peak in my graph, and the arrow is also on the same... [click a timestamp on peers' affect]... I should have read (about this video content) before so I don't associate with this (peers') low-intensity peak."* Interview quotes suggested that all *Locative* participants found the peers' cues to be very *"intuitive"* and *"heuristic"* to make sense of and found these cues useful to assist them in interpreting their own graph. At the same time, *Locative* participants reported only a few possible alignments regarding their own and peers' spikes and felt discouraged to reflect.

***Temporal* cues associated with recalling video content and building shared knowledge.** We found a strong and positive correlation between clicking peers' affects and verbalizing *(dis)agreeing with peers* ($r = .894$, $p < .001$), with the power analysis resulting in a value equal to 0.92. We also found a strong and positive correlation between clicking peers' affects and verbalizing *building shared knowledge* ($r = .820$, $p < .001$), with the power analysis resulting in a value equal to 0.90. The *Temporal* participants compared their own graph to peers' affects around timestamps chronologically (from left to right on the interface). Upon explaining their own learning affects and becoming more aware of themselves by *recalling content to explain own affects*, participants were able to verbalize their awareness of the differences between their own and peers' learning affects in order to *build shared knowledge*.

During think-aloud, *Temporal* Participants expressed a strong interest in reflecting on the *whole* video content and segmenting the video into several sections to reflect on. More specifically, words describing temporal aspects of the video instead of clicking on their own graph, e.g., *"part, section"*, were used 33 times, compared to less than ten times in total by the other two groups. An example quote from T10 is, "*I already mentioned that there were at least two standout sections in this presenter that I found visually engaging and humorous.*" Especially, they need to visually segment their own graph into positive, negative, and neutral sections based on graph color while recalling video content. An example think-aloud quote from T03 confirmed the segmenting process: *"I think the biggest difference (between me and others), hmmm, is at the last section. It's where the video talks about the mechanism of AR and some details of its visualization."* When reviewing their peers' cues, individuals closely read peers' comments to recall video content when their own affect graph shows an equal mix of warm and cool colors, making it difficult to differentiate between positive and negative affects and further comparison. Most *Temporal* participants said during the interviews that they compared their own affects with peers' and reviewed their own graph at nearby sections to *"contextualize own and peers' affects"* and *"find similarities/differences,"* which is cognitive-demanding.

The *Temporal* participants showed curiosity about "Tied" content. *Temporal* participants perceived *Temporal* cues to cover the whole video by assigning meanings to blank segments in Figure 1 as neither 'agreed negative' nor 'agreed positive,' with 50% expressed positively and 50% expressed negatively. Such 50-50 contradiction indicates the content to be controversial and worth of discussion. T08 explained how such segments intrigued her to read peers' comments during the interview: *"I am so excited to read more peers' comments on a segment where people don't have an agreed positive content, not just one, conversations are even better. I love to see how people hate each other! ... Seriously, that's something you can hardly see in person, but let you know how different people think differently."*

Summary: First, by coding participants' think-aloud data, we identified three types of meta-cognition processes when reflecting on peers' cues: *(dis)agreement with peers affects*, *recalling content to explain own affects*, and *building shared knowledge*. *Temporal* participants verbalized significantly more on *building shared knowledge* than participants in the two other groups, as shown in Figure 5. Furthermore, the survey results showed that participants also perceived the two proposed affect cues as significantly more useful in supporting their meta-cognition.

Additionally, *Temporal* and *Locative* cues had different effects on the types of meta-cognition processes conducted during think-aloud. Specifically, clicking *Temporal* cues (aggregating peers' positive/negative emotions) encouraged participants *recalling content to explain own affects* and further verbalize on *building shared knowledge*, as shown in Table 1. *Temporal* participants segmented their own graphs based on positivities and intended to reflect on the whole video, resulting in a cognitively engaging process.

5.2 Participants' Perceptions of Sharing Own Affects with Peers (RQ2)

We compared participants' survey scores in the three groups to examine the effect of the proposed designs on participants' trust in the tool, awareness of the system limitations, and perceived privacy concerns. The interview results explained some effects.

5.2.1 Potential Factors Mitigating the Trust in AI-Generated Peers' Affects. Survey results showed that participants in all three groups agreed that peers' cues increased trust, and disagreed that peers' cues decreased trust. Additionally, participants reflected on the two AI-generated affects and identified factors that mitigated their trust in the system.

When asked to rate agreement in "I think navigating peers' cues increases my trust in the tool then only seeing my own emotion," participants in all three groups found *MirrorUs* increased their trust. No effect of group membership was found on increased trust using ANOVA testing. *Locative* participant (M=4.9, SD=1.2), *Temporal* participant(M=4.2, SD=1.4), and participants from the control group (M=4.8, SD=1.6) all perceived that navigating peers' cues improved their trust in using the tool. Interview results suggest that while navigating peers' cues, participants' trust increased as the peers' cues led them to believe many other users were also using *MirrorUs* and created a sense of social presence.

When asked "I think navigating peers' cues decreases my trust in the tool then only seeing my own emotion," participants in all three groups disagreed that navigating peers' cues decreased trust. Compared to participants from the control group that disagreed with navigating peers' decreasing trust, participants in the two treatment groups were significantly more neutral towards decreased trust. Specifically, ANOVA and Post-hoc comparisons using the Tukey HSD test showed *Locative* participants ($M = 2.9$, $SD = 1.1$) ($p < .01$) and *Temporal* participants ($M = 2.9$, $SD = 0.8$) ($p < .05$) reported significantly higher value than that of the control group ($M = 1.8$, $SD = 0.7$, $F(2,39) = 6.5$, $p < .01$). There was no significant difference between the two treatment group. The resulting power is 0.82.

Compared to the control group that had shared agreement towards decreasing trust, participants in the two treatment groups had more diverse views, and a few strongly agreed that navigating peers' cues decreased trust. Participants in the two treatment groups 1) suggested more transparency in how text comments are selected to motivate sharing, and 2) complained that the accuracy of the AER model needs to be improved.

The majority of participants in the two treatment groups expressed that the system only displayed one peer's comment for each timestamp, which caused them to wonder if the system was hiding any comments and how it selected the comments. For example, L10 said *"I wonder why (the system) only displays one text comment (for each arrow showing common emotion intensity peak). I bet you have collected many comments, as I provided five myself. How did you select which one to display? Are you hiding some comments? ... I want more transparency. Yeah, more transparency, I wouldn't say I don't trust it... I would like more transparency in order to motivate me to share."* The quote indicates how the system displayed peers' comments triggering participants to start thinking about how their own comment is processed by the system. Some participants questioned whether the AER data displayed was aggregated from multiple participants or if only the affect of the comment contributor was given higher weight when they saw only one peer's text comments for each timestamp.

Another theme is that when navigating AI-generated peers' cues, the inaccuracy in AER made participants in two treatment groups less willing to trust or share their own AER data. By navigating peers' affects, they got a sense of how their own affects data would be processed by the system (e.g. extracting AER spikes to generate aggregated *emotion spikes*) and displayed to peers. However, they don't perceive such processes to represent their own effects accurately and fairly, leading

to a decrease in trust. For example, L14, as shown in Figure 2, said: *"My graph does not display any intense spikes, which could be attributed to the fact that I may express emotions differently than other people. To ensure better accuracy and inclusion, the system should incorporate positive/negative qualities that more accurately reflect my emotions when compared to other users. This will make the AI and comparison more trustworthy and appealing, likely leading to more people utilizing the system."*

5.2.2 AI-Generated Peers' Affects Improving the Awareness of Learning Affects, however, with Limitations. Survey results showed that *Locative* participants rated the highest level of agreement (M= 5.7, SD=0.8) when asked to assess the statement: "I think navigating peers' cues help me notice the actual evidence behind the tool than only seeing my own emotion." The results of the *Locative* study indicate that peers' cues were the most effective in helping participants to become aware of their own and their peers' emotions, as well as to recognize the system's constraints that prevented them from noticing affects. *Temporal* participants (M=3.8, SD=1.3) and those from the control group (M=4.4, SD=1.7) were neutral towards awareness of the tool limitations. ANOVA and Post-hoc comparisons using the Tukey HSD test indicated a significant difference at $p < 0.01$ level between *Locative* participants and participants from the two other groups. The resulting power is 0.90.

Data from interviews revealed that the system's main limitation was its lack of capability to give sufficient explanations that reflect the views of varied individuals. The insufficient explanation hindered participants' ability to comprehend and empathize with peers' affects. L13's quote explains the expectation for "diversity visibility": *"I was actually expecting to see more diverse opinions and more text comments, like in a classroom discussion. The majority is not always right, and someone and the minority might offer thought-provoking perspectives. And sometimes, people who seem to be different from the majority are the ones that really need resonating and peers' emotional support!"* The quotes suggest participants need semantically rich explanations to understand AI-generated peers' cues, especially to make sense of different causes of affects and resonate with them.

Eight participants from the two treatment groups suggested that the system should display text comments for both majority and minority AER-based affects. T10 gave an explanation summarising multiple reasons for this, *"People can feel the same way for different reasons, or people should have the right to express differently, and we as viewers should also be given the opportunities to see from different points of view. The current design could cause people not to have individual thoughts and stop those who feel different from others from expressing themselves. And in class, knowing that someone else also has negative emotions makes learners more confident in asking questions."* Her explanation, from both the contributor and consumer perspectives, revealed the value of increasing the visibility of different viewpoints. Showing fewer peers inputted text comments makes AI-generated peers' affects less authentic and discouraging for learners to share their own text comments. Our current interview themes could not explain the differences between *Locative* participants and *Temporal* participants in different awareness of limitation, and the discussion section revealed possible explanations.

5.2.3 AI-Generated Peers' Affects Raising More Privacy Concerns. When asked in the survey "I think navigating peers' cues increases my privacy concerns more than only seeing my own emotion," participants in the two treatment groups perceived significantly higher privacy concerns. On average, *Locative* participants had a slight increase in privacy concerns (M = 4.2, SD = 1.9), Temporal participants were neutral towards increased privacy concerns (M = 3.5, SD = 2.3), while participants in the control group disagree with increased privacy concern (M = 1.6, SD = 0.7). ANOVA found a significant group effect ($F(2, 39) = 7.918$, $p < .01$) on increased privacy concerns and Post-hoc comparisons using the Tukey HSD test showed that the mean score for *Locative* participants ($p < .01$) and *Temporal* participants ($p < .01$) was significantly higher than that of participants from the control group. The resulting power is 0.97.

According to the interview, participants were more concerned about how their facial images were processed by the system. For example, the peers' cues made the participants wonder whether their facial images were stored elsewhere, instead of in their own local computers, as T07 shared, *" After watching the video and seeing my own results, I was interested to see how my data was being processed and sent to where both my self-report and facial data were being stored. I was a bit concerned about the privacy of my data, but I appreciated that you asked us to see the peers' cues after viewing the results. If I had been informed earlier about how my data was being shared with others, I would likely have been less expressive with my facial expressions because I was worried about it (privacy issue) ".* Participants suggested that the system should only display aggregated affects of peers. They also suggested the system notify the users whether their facial images were uploaded to a cloud server whenever the system displayed information related to facial information, such that users would not be surprised to see the cues.

Summary: The results indicated that participants in the two treatment groups felt a lack of transparency in the display of peers' text comments and the accuracy of the AER model, leading to mitigated trust. Interviews revealed that more semantically rich comments could help participants better resonate and comprehend their peers' affects. Additionally, the two AI-generated peers' affect cues caused participants to be significantly more concerned about privacy than the control group.

6 DISCUSSION

6.1 Social Meta-Cognition through AER-based Peers' Affects

Through need-finding, we proposed two types of cues by AI-generated AER-based peers' affects. Specifically, the *Locative* cues aggregated common *emotion spikes* among peer learners, and the *Temporal* cues aggregated positive and negative *dominant emotions* among peers. According to participants' use of the cues and their survey input, both designs were found to be significantly more useful for meta-cognition and easier for understanding their own affects than without providing the affect cues (in the control group). The AER solution enabled tracking affects continuously and visualizing them in segments along the video timelines at a low cost.

This study bridges the knowledge gap in the relationship between regulation and meta-cognition pointed out by recent literature review work in online learning [13] by finding reflecting on peers' affects aggregated from AER can support an individual's development of meta-cognition in video-based learning. Think-aloud data in RQ1 showed participants verbalized both meta-cognition from an individual and social perspective using AER-based peer affects and we further explain it below. Expanding on [35]'s work on AER, we empirically found that aggregating the two distinct metrics of AER, *emotion spikes* and *dominant emotions*, lead to different effects on meta-cognition, as evidenced by RQ1. As shown in Figure 5, one of the two proposed peers' affects cues, *Temporal* cues, elicited significantly more quantity on *building shared knowledge* than the two other groups. This indicates that reflecting on peers' common positive/negative affects can promote the development of *social meta-cognition*, where group members' questions, evaluations, repetitions, and elaborations help individuals see their limitations, build shared knowledge, and expand their understandings [16, 30]. Clicking *Temporal* cues elicited more comparison opportunities between self and peers, leading to a higher quantity of meta-cognitive processes - *recalling content to explain own affects*, and *building shared knowledge* as shown in Table 1.

However, RQ2 results suggest that peers' AER cannot replace peers' semantic narrative of affects. Moreover, for all three groups, the tool displayed peers' text comments along with the AER data, which can outperform reflection with its own affects [15] by supporting new meta-cognition

processes in RQ1 - *(dis)agreement with peers affects* and *building shared knowledge*. Further research should better examine the complementary relationship between learner-inputted text and AER-based data. Future research can further explore how different types of affects inter-relate and how to collect more genuine data in a more natural and efficient way, for example, by using AER to detect learning moments and remind learners to provide comments to share with peers.

6.2 Design Implications

We provide design implications to support effective reflection with peers' affects informed by RQ1 and discuss future design goals to address participants' concerns in sharing their own affects RQ2.

6.2.1 Temporal and Locative Cues at Different Learning Stages (RQ1). From the participants' think-aloud, survey, and interview data obtained in RQ1, we discovered that *Locative* cues support fast and heuristic understanding of affect, i.e., extracting *emotion spikes* in their own graph, while *Temporal* cues require a contextualized understanding to recall video context and explain own affects. We provide implications for the use of both cues in pre-, in-, and post-learning. For in-class learning, where instructions are usually provided, real-time *Locative* cues can be presented with learners' own emotion spikes to quickly capture learners' attention when mind-wandering is detected while minimizing the use of cognitive resources. For post-class learning, where reflections are common [42, 49, 68], *Temporal* cues with own dominant emotions can be used to facilitate diverse types of reflections, specifically recalling video content and explaining how/why learners' agree with peers. To make the *Temporal* cues more effective in facilitating successful recalls, the system can extract video content to help with recall, e.g., by using NLP to divide the video into segments based on the learning material. For pre-class, effort planning and cognitive engagement take place [42], *Locative* cues can be employed to increase motivation (e.g., system displaying viewers have expressed excitement in three knowledge points) together with *Temporal* cues (e.g., system displaying viewers expressed negative and found 20% of video boring) to give learners an overview of the class and support task analysis.

6.2.2 Promises in Designing Peers' Affect Cues for Reflections (RQ2). Our RQ2 found that the two peers' affect cues improved trust in the system and elicited significantly greater privacy concerns than participants from the control group when sharing AER-based affects with peers. Our research provides insights for future directions on engaging participants in genuinely shared modes of regulation.

Before beginning active learning, addressing privacy concerns is essential. It is not surprising that sharing AER increases privacy concerns in RQ2. Participants' preferences in need-finding sessions to aggregating affects anonymously did not fully address privacy concerns. It confirms prior research [47], that found participants consider affects data to be personal and wanted AER algorithms to run locally. Despite some participants saying the concerns helped them to be more conscious of using technology in post-learning reflections, they perceive the concerns to be distractive while learning. To allow learners to focus on the learning task designs that share AER should take into account these privacy concerns beforehand.

Addressing limitations in AI by providing diverse and semantically-rich explanations. Participants from both *Temporal* and *Contextual* groups suggested displaying contradicting comments to better understand and resonate with peers' affective states in RQ2. This aligns with the suggestion by *Temporal* participants under RQ1 to allow learners to see "tied" content where a similar amount of peers are negative and positive. Participants in both two groups also mentioned more peers' comments that contradict each other's sentiments, as well as better designs that can show multiple comments simultaneously, such as *danmauku*. Danmaku comments are overlaid on the screen of

videos without displaying users' information and have been found to significantly promote user participation compared to traditional linear comment sections [63, 64].

Improving accessibility by enabling social connectivity. As mentioned in the trust section under RQ2, social presence introduced by peers' cues was appreciated and was associated with increased trust in all three groups. Peers' affects can also provide opportunities to increase accessibility and promote inclusive learning among deaf and hard-of-hearing students, who often rely on self-teaching strategies in classrooms and online learning environments, and are commonly isolated as facial expressions are often inaccessible to them in front-facing classrooms [52]. Considering this, peers' affect cues and comments should be adapted to different learners' communities and probe social connectedness opportunities between them, such as collaborative learning to augment video-based learning captioning [7]. For example, AER may also be utilized to detect when deaf and hard-of-hearing students show confusion and frustration and provide social support accordingly.

6.3 Expanding an XAI Design Framework towards Social Transparency in AI Systems

A new design framework for AI called *Social Transparency*, has recently been introduced. This socio-technically informed perspective takes into account the socio-organizational context in order to explain AI-mediated decision-making in industry settings. This framework includes the "4W" design categories: *What* - Action taken on AI decision outcome, *Why* - Comments with rationale justifying the decision, *Who*, and *When*, and has been found to calibrate trust in AI, improve decision-making, and cultivate holistic explainability [25]. Building on this work, we make an implication by proposing the addition of a new design category to the 4W framework: *How*. Our findings suggested that participants intended to understand how their AI-generated affect is generated. This led to participants' need for more transparency in AI-based systems and comments on AER algorithm accuracy. Our findings also showed that peers' text comments (explaining how the peers' cues were generated) helped participants reflect on how their own data was processed by the AI-based system. This can be explained as a *Social Mirror* process, which is defined as predicting different emergent levels of consciousness, together with their attendant perceptions of need, according to what the social environment can reflect [62].

As such, we suggest that *explaining how the AI results are generated* should be included to inform and empower end-users to take informed and accountable actions. Further research should investigate how the *How + 4W* framework may be used to enhance Social Transparency in AI-based systems. The *How* focuses on the purpose of end-user reflections and could be insightful for other settings that aim to improve self-consciousness, such as self-tracking of mental/physical health [5, 38].

6.4 Limitations and Future Work

There are a few limitations in our study. First, we only used AER via facial movements to recognize the emotion of the learners. There are also other AER technology, such as body movements [65], speech [59], and brain activities [1]. Second, our study only included current college students, and the results may not apply to all populations. More studies should be conducted on a more diverse population. Third, in our system, the common *emotion spikes* and *dominant emotions* were manually labeled by the participants and then aggregated by researchers. Future works may examine how to automatically recognize emotions considering grid search to choose the more preferred time granularity [35].

7 CONCLUSION

Recent research has examined the effects of awareness of peers' affect on learners' meta-cognition in online learning, however, few studies have explored the implications of using Automatic Emotion

Recognition (AER) technology to track and visualize peers' cues in order to facilitate meta-cognition. To investigate this, two types of AI-generated cues were proposed: *Locative* cues (peers' common emotion spikes) and *Temporal* cues (peers' common positive/negative dominant emotions). Think-aloud, interview, and survey methods were used to evaluate the effects of these cues on learner meta-cognition, as well as the potential ethical concerns of sharing personal affective data. Results of this study suggest that peers' AER-based data can be advantageous for learners' meta-cognition, however, potential issues with privacy and trust should be taken into consideration. Results found that the two cues supported different meta-cognitive processes, and *Temporal* cues supported significantly more explanations of how and why learners (dis)agreed with and/or (fail to) understand peers' affects than the two other cues. While the learners perceived the benefits of using AI-generated peers' cues on improving the awareness of their own learning affects, they also looked for more peers' narratives that can explain the AI results.

8 ACKNOWLEDGEMENT

This material is based upon work supported by the National Science Foundation under Grant No. 2119589. This material is also based upon work supported by the AI Research Institutes program by the National Science Foundation and the Institute of Education Sciences, U.S. Department of Education through Award No. 2229873 - AI Institute for Transforming Education for Children with Speech and Language Processing Challenges. Any opinions, findings and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation, the Institute of Education Sciences, or the U.S. Department of Education. Any opinions, findings, conclusions, or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation, the Institute of Education Sciences, or the U.S. Department of Education.

REFERENCES

[1] Pascal Ackermann, Christian Kohlschein, Jó Agila Bitsch, Klaus Wehrle, and Sabina Jeschke. 2016. EEG-based automatic emotion recognition: Feature extraction, selection and classification methods. In *2016 IEEE 18th international conference on e-health networking, applications and services (Healthcom)*. IEEE, 1–6.

[2] Muhammad Afzaal, Jalal Nouri, Aayesha Zia, Panagiotis Papapetrou, Uno Fors, Yongchao Wu, Xiu Li, and Rebecka Weegar. 2021. Explainable AI for data-driven feedback and intelligent action recommendations to support students self-regulation. *Frontiers in Artificial Intelligence* 4 (2021).

[3] Roger Azevedo, Michelle Taub, Nicholas V Mudrick, Garrett C Millar, Amanda E Bradbury, and Megan J Price. 2017. Using data visualizations to foster emotion regulation during self-regulated learning with advanced learning technologies. In *Informational environments*. Springer, 225–247.

[4] Lisa Feldman Barrett, Ralph Adolphs, Stacy Marsella, Aleix M Martinez, and Seth D Pollak. 2019. Emotional expressions reconsidered: Challenges to inferring emotion from human facial movements. *Psychological science in the public interest* 20, 1 (2019), 1–68.

[5] Marit Bentvelzen, Paweł W Woźniak, Pia SF Herbes, Evropi Stefanidi, and Jasmin Niess. 2022. Revisiting Reflection in HCI: Four Design Resources for Technologies that Support Reflection. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 6, 1 (2022), 1–27.

[6] Patricia EG Bestelmeyer, Sonja A Kotz, and Pascal Belin. 2017. Effects of emotional valence and arousal on the voice perception network. *Social cognitive and affective neuroscience* 12, 8 (2017), 1351–1358.

[7] Bhavya Bhavya, Si Chen, Zhilin Zhang, Wenting Li, Chengxiang Zhai, Lawrence Angrave, and Yun Huang. 2022. Exploring collaborative caption editing to augment video-based learning. *Educational technology research and development* 70, 5 (2022), 1755–1779.

[8] Virginia Braun and Victoria Clarke. 2006. Using thematic analysis in psychology. *Qualitative research in psychology* 3, 2 (2006), 77–101.

[9] Michael Brown. 2020. Seeing students at scale: How faculty in large lecture courses act upon learning analytics dashboard data. *Teaching in Higher Education* 25, 4 (2020), 384–400.

[10] Zana Buçinca, Maja Barbara Malaya, and Krzysztof Z Gajos. 2021. To trust or to think: cognitive forcing functions can reduce overreliance on AI in AI-assisted decision-making. *Proceedings of the ACM on Human-Computer Interaction* 5,

CSCW1 (2021), 1–21.
[11] Rafael A Calvo and Sidney K D'Mello. 2011. *New perspectives on affect and learning technologies*. Vol. 3. Springer Science & Business Media.
[12] Stephane Champely. 2021. *R package "pwr"*. http://cran.r-project.org/web/packages/pwr/ R package version 1.3.
[13] Carol KK Chan. 2012. Co-regulation of learning in computer-supported collaborative learning environments: A discussion. *Metacognition and learning* 7, 1 (2012), 63–73.
[14] Si Chen, Desirée Kirst, Qi Wang, and Yun Huang. 2023. Exploring Think-aloud Method with Deaf and Hard of Hearing College Students. In *Proceedings of the 2023 ACM Designing Interactive Systems Conference*. 1757–1772.
[15] Si Chen, Yixin Liu, Risheng Lu, Yuqian Zhou, Yi-Chieh Lee, and Yun Huang. 2022. " Mirror, Mirror, on the Wall"-Promoting Self-Regulated Learning using Affective States Recognition via Facial Movements. In *Designing Interactive Systems Conference*. 1300–1314.
[16] Ming Chiu and S.W. Kuo. 2009. Social metacognition in groups: Benefits, difficulties, learning, and teaching. *Metacognition: New Research Developments* (01 2009), 117–136.
[17] Eun Kyoung Choe, Bongshin Lee, Haining Zhu, Nathalie Henry Riche, and Dominikus Baur. 2017. Understanding self-reflection: how people reflect on personal data through visual data exploration. In *Proceedings of the 11th EAI International Conference on Pervasive Computing Technologies for Healthcare*. 173–182.
[18] Cristina Conati, Oswald Barral, Vanessa Putnam, and Lea Rieger. 2021. Toward personalized XAI: A case study in intelligent tutoring systems. *Artificial Intelligence* 298 (2021), 103503.
[19] Popa Daniela. 2015. The relationship between self-regulation, motivation and performance at secondary school students. *Procedia-Social and Behavioral Sciences* 191 (2015), 2549–2553.
[20] Didan Deng, Zhaokang Chen, and Bertram E Shi. [n.d.]. Multitask Emotion Recognition with Incomplete Labels. In *2020 15th IEEE International Conference on Automatic Face and Gesture Recognition (FG 2020)(FG)*. IEEE Computer Society, 828–835.
[21] Benedict Du Boulay, Katerina Avramides, Rosemary Luckin, Erika Martínez-Mirón, Genaro Rebolledo Méndez, and Amanda Carr. 2010. Towards systems that care: a conceptual framework based on motivation, metacognition and affect. *International Journal of Artificial Intelligence in Education* 20, 3 (2010), 197–229.
[22] Sidney D'Mello, Arvid Kappas, and Jonathan Gratch. 2018. The affective computing approach to affect measurement. *Emotion Review* 10, 2 (2018), 174–183.
[23] Anastasia Efklides. 2006. Metacognition and affect: What can metacognitive experiences tell us about the learning process? *Educational research review* 1, 1 (2006), 3–14.
[24] Anastasia Efklides, Bennett L Schwartz, and Victoria Brown. 2017. Motivation and affect in self-regulated learning: does metacognition play a role? In *Handbook of self-regulation of learning and performance*. Routledge, 64–82.
[25] Upol Ehsan, Q Vera Liao, Michael Muller, Mark O Riedl, and Justin D Weisz. 2021. Expanding explainability: Towards social transparency in AI systems. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. 1–19.
[26] Mohamed Ez-Zaouia, Aurélien Tabard, and Elise Lavoué. 2020. Emodash: A dashboard supporting retrospective awareness of emotions in online learning. *International Journal of Human-Computer Studies* 139 (2020), 102411.
[27] Jeffrey M Girard and Aidan G C Wright. 2018. DARMA: Software for dual axis rating and media annotation. *Behavior research methods* 50, 3 (2018), 902–909.
[28] Gabriel Grill and Nazanin Andalibi. 2022. Attitudes and Folk Theories of Data Subjects on Transparency and Accuracy in Emotion Recognition. *Proceedings of the ACM on Human-Computer Interaction* 6, CSCW1 (2022), 1–35.
[29] Luke Guerdan, Alex Raymond, and Hatice Gunes. 2021. Toward affective XAI: facial affect analysis for understanding explainable human-ai interactions. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*. 3796–3805.
[30] Cindy E Hmelo-Silver and Howard S Barrows. 2008. Facilitating collaborative knowledge building. *Cognition and instruction* 26, 1 (2008), 48–94.
[31] Wayne Holmes, Kaska Porayska-Pomsta, Ken Holstein, Emma Sutherland, Toby Baker, Simon Buckingham Shum, Olga C Santos, Mercedes T Rodrigo, Mutlu Cukurova, Ig Ibert Bittencourt, et al. 2021. Ethics of AI in education: Towards a community-wide framework. *International Journal of Artificial Intelligence in Education* (2021), 1–23.
[32] Lun-Kai Hsu, Wen-Sheng Tseng, Li-Wei Kang, and Yu-Chiang Frank Wang. 2013. Seeing through the expression: Bridging the gap between expression and emotion recognition. In *2013 IEEE International Conference on Multimedia and Expo (ICME)*. IEEE, 1–6.
[33] M Sazzad Hussain, Omar AlZoubi, Rafael A Calvo, and Sidney K D'Mello. 2011. Affect detection from multichannel physiology during learning sessions with AutoTutor. In *International Conference on Artificial Intelligence in Education*. Springer, 131–138.
[34] Gwo-Jen Hwang, Sheng-Yuan Wang, and Chiu-Lin Lai. 2021. Effects of a social regulation-based online learning framework on students' learning achievements and behaviors in mathematics. *Computers & Education* 160 (2021),

104031.
[35] Harmanpreet Kaur, Daniel McDuff, Alex C Williams, Jaime Teevan, and Shamsi T Iqbal. 2022. "I Didn't Know I Looked Angry": Characterizing Observed Emotion and Reported Affect at Work. In *CHI Conference on Human Factors in Computing Systems*. 1–18.
[36] Hassan Khosravi, Simon Buckingham Shum, Guanliang Chen, Cristina Conati, Yi-Shan Tsai, Judy Kay, Simon Knight, Roberto Martinez-Maldonado, Shazia Sadiq, and Dragan Gašević. 2022. Explainable artificial intelligence in education. *Computers and Education: Artificial Intelligence* 3 (2022), 100074.
[37] Joongyum Kim, Taesik Gong, Kyungsik Han, Juho Kim, JeongGil Ko, and Sung-Ju Lee. 2020. Messaging beyond texts with real-time image suggestions. In *22nd International Conference on Human-Computer Interaction with Mobile Devices and Services*. 1–12.
[38] Rafal Kocielnik, Lillian Xiao, Daniel Avrahami, and Gary Hsieh. 2018. Reflection Companion: A Conversational System for Engaging Users in Reflection on Physical Activity. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 2, 2, Article 70 (jul 2018), 26 pages. https://doi.org/10.1145/3214273
[39] Barry Kort, Rob Reilly, and Rosalind W Picard. 2001. External representation of learning process and domain knowledge: Affective state as a determinate of its structure and function. In *Workshop on Artificial Intelligence in Education (AI-ED 2001), San Antonio,(May 2001)*. 64–69.
[40] Kostiantyn Kucher, Carita Paradis, and Andreas Kerren. 2018. The state of the art in sentiment visualization. In *Computer Graphics Forum*, Vol. 37. Wiley Online Library, 71–96.
[41] Haotian Li, Min Xu, Yong Wang, Huan Wei, and Huamin Qu. 2021. A visual analytics approach to facilitate the proctoring of online exams. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. 1–17.
[42] Shan Li and Susanne P Lajoie. 2021. Cognitive engagement in self-regulated learning: An integrative model. *European Journal of Psychology of Education* (2021), 1–20.
[43] Q Vera Liao and Kush R Varshney. 2021. Human-centered explainable ai (xai): From algorithms to user experiences. *arXiv preprint arXiv:2110.10790* (2021).
[44] 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*. 1–25.
[45] Prashan Madumal, Ronal Singh, Joshua Newn, and Frank Vetere. 2018. Interaction design for explainable AI: workshop proposal. In *Proceedings of the 30th Australian Conference on Computer-Human Interaction*. 607–608.
[46] Michael E Martinez. 2006. What is metacognition? *Phi delta kappan* 87, 9 (2006), 696–699.
[47] 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.
[48] Omid Noroozi, Héctor J Pijeira-Díaz, Marta Sobocinski, Muhterem Dindar, Sanna Järvelä, and Paul A Kirschner. 2020. Multimodal data indicators for capturing cognitive, motivational, and emotional learning processes: A systematic literature review. *Education and Information Technologies* 25 (2020), 5499–5547.
[49] Ernesto Panadero. 2017. A review of self-regulated learning: Six models and four directions for research. *Frontiers in psychology* 8 (2017), 422.
[50] Monica Pereira and Kate Hone. 2021. Communication skills training intervention based on automated recognition of nonverbal signals. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. 1–14.
[51] Michael J Reiss. 2021. The use of AI in education: Practicalities and ethical considerations. *London Review of Education* (2021).
[52] John TE Richardson, Gary L Long, and Susan B Foster. 2004. Academic engagement in students with a hearing loss in distance education. *Journal of Deaf Studies and Deaf Education* 9, 1 (2004), 68–85.
[53] Kat Roemmich and Nazanin Andalibi. 2021. Data subjects' conceptualizations of and attitudes toward automatic emotion recognition-enabled wellbeing interventions on social media. *Proceedings of the ACM on Human-Computer Interaction* 5, CSCW2 (2021), 1–34.
[54] James A Russell. 1980. A circumplex model of affect. *Journal of personality and social psychology* 39, 6 (1980), 1161.
[55] Maryam Salehomoum. 2023. Think-Aloud: Effect on Adolescent Deaf Students' Use of Reading Comprehension Strategies. *The Journal of Deaf Studies and Deaf Education* 28, 1 (2023), 99–114.
[56] Samiha Samrose, Daniel McDuff, Robert Sim, Jina Suh, Kael Rowan, Javier Hernandez, Sean Rintel, Kevin Moynihan, and Mary Czerwinski. 2021. MeetingCoach: An Intelligent Dashboard for Supporting Effective & Inclusive Meetings. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. 1–13.
[57] Tanya Sharma, Manoj Diwakar, and Chandrakala Arya. 2022. A systematic review on emotion recognition by using machine learning approaches. In *AIP Conference Proceedings*, Vol. 2481. AIP Publishing LLC, 020045.
[58] Kanchan S Vaidya, Pradeep M Patil, and Mukil Alagirisamy. 2022. A Review of Human Emotion Recognition. *Specialusis Ugdymas* 1, 43 (2022), 1423–1451.

[59] Thurid Vogt and Elisabeth André. 2005. Comparing feature sets for acted and spontaneous speech in view of automatic emotion recognition. In *2005 IEEE International Conference on Multimedia and Expo*. IEEE, 474–477.
[60] Guanfeng Wang, Chen Gong, and Shuxia Wang. 2022. A Review of Automatic Detection of Learner States in Four Typical Learning Scenarios. In *Adaptive Instructional Systems*, Robert A. Sottilare and Jessica Schwarz (Eds.). Springer International Publishing, Cham, 53–72.
[61] Adrian Wells and Sam Cartwright-Hatton. 2004. A short form of the metacognitions questionnaire: properties of the MCQ-30. *Behaviour research and therapy* 42, 4 (2004), 385–396.
[62] Charles Whitehead. 2001. Social mirrors and shared experiential worlds. *Journal of consciousness studies* 8, 4 (2001), 3–36.
[63] Qunfang Wu, Yisi Sang, and Yun Huang. 2019. Danmaku: a new paradigm of social interaction via online videos. *ACM Transactions on Social Computing* 2, 2 (2019), 1–24.
[64] Qunfang Wu, Yisi Sang, Shan Zhang, and Yun Huang. 2018. Danmaku vs. forum comments: understanding user participation and knowledge sharing in online videos. In *Proceedings of the 2018 ACM conference on supporting groupwork*. 209–218.
[65] Haris Zacharatos, Christos Gatzoulis, and Yiorgos L Chrysanthou. 2014. Automatic emotion recognition based on body movement analysis: a survey. *IEEE computer graphics and applications* 34, 6 (2014), 35–45.
[66] Wei Zhan, Jyhwen Wang, Manoj Vanajakumari, and Michael D Johnson. 2018. Creating a High Impact Learning Environment for Engineering Technology Students. *Advances in Engineering Education* 6, 3 (2018), n3.
[67] Ru Zhao, Vivian Li, Hugo Barbosa, Gourab Ghoshal, and Mohammed Ehsan Hoque. 2017. Semi-automated 8 collaborative online training module for improving communication skills. *Proceedings of the ACM on interactive, mobile, wearable and ubiquitous technologies* 1, 2 (2017), 1–20.
[68] Barry J Zimmerman. 1990. Self-regulated learning and academic achievement: An overview. *Educational psychologist* 25, 1 (1990), 3–17.
[69] Barry J Zimmerman. 2013. From cognitive modeling to self-regulation: A social cognitive career path. *Educational psychologist* 48, 3 (2013), 135–147.

APPENDICES

A.1 SURVEY QUESTIONS

We list all our survey questions below. All Questions are in 1-7 Likert scale with 1 as strongly disagree, 2 as disagree, 3 as moderately disagree, 4 as neutral, 5 as moderately agree, 6 as agree, and 7 as strongly agree.

(1) Usefulness for Metacognition
  (a) Cognitive Confidence
    - This tool can help me improve trust in my memory in video-based learning
    - This tool can help me overcome poor memory in video-based learning
    - This tool can help me increase confidence in my memory for actions in video-based learning
    - This tool can help me increase confidence in my memory for the learning content in video-based learning
    - This tool can help me avoid my memory from misleading me at times in video-based learning
  (b) Cognitive Self-Consciousness
    - This tool can help me be constantly aware of my thinking in video-based learning
    - This tool can help me think a lot about my thoughts in video-based learning
    - This tool can help me constantly examine my thoughts in video-based learning
    - This tool can help me monitor my thoughts in video-based learning
    - This tool can help me be aware of the way my mind works when I am thinking through a problem in video-based learning
(2) Cognitive Demand
  - I found reflecting on my own emotion difficult.
  - I found reflecting on peers' cues difficult.
(3) Privacy Concerns
  - I think navigating peers' emotions/cues increases my privacy concerns than only seeing my own emotion
(4) Awareness in Limitations of data, bias, and representation
  - I think navigating peers' emotions/cues help me notice the actual evidence behind the tool than only seeing my own emotion
(5) Trust
  - I think navigating peers' emotions/cues increases my trust in the tool than only seeing my own emotion
  - I think navigating peers' emotions/cues decreases my trust in the tool than only seeing my own emotion

A.2 SUPPORTIVE IMAGES FOR NEED-FINDING STUDIES

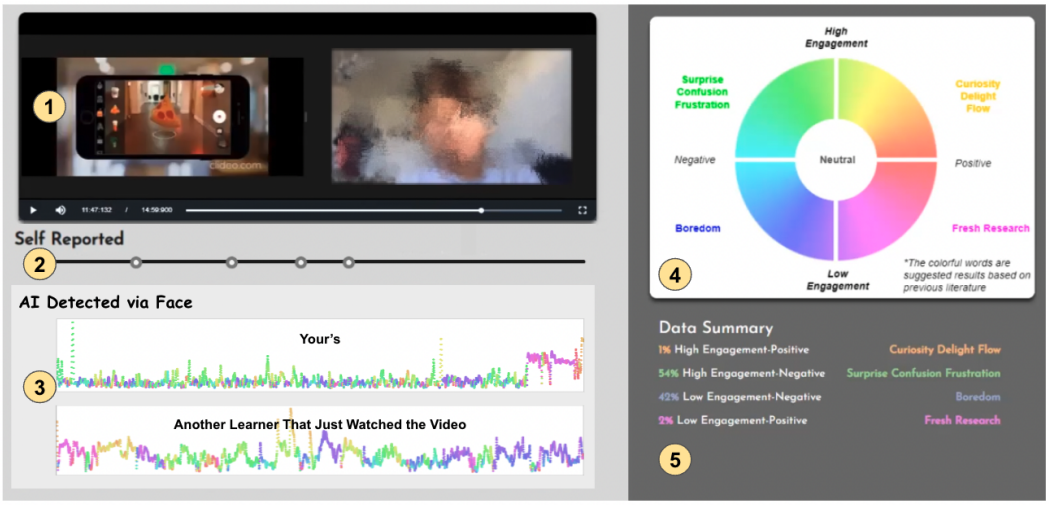

Fig. A1. Post-Learning Interface used in Round I. Participants could navigate the five areas illustrated above: (1) playback of the video along with the participants' selfie captured when watching the same moment of the video, (2) tagged self-reported moments of affective states, (2) time-based AI results of the participants and the other learner's affective states via facial movements, (4) a legend for interpreting time-based AI results, and (5) the summary of AI results for the participants' data.

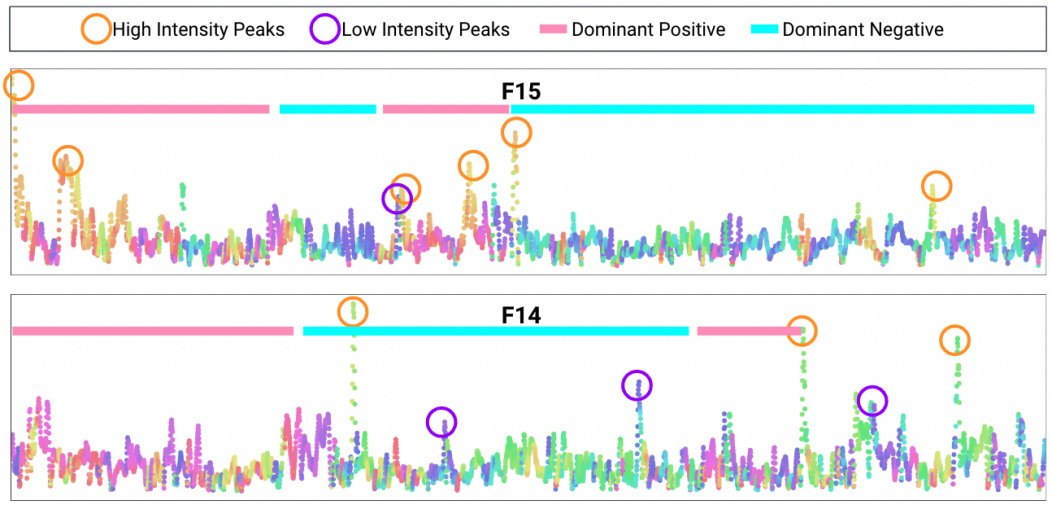

Fig. A2. Two examples of participants self-labeled emotion intensity graph in Round II of the study. A total of ten participants from Round I labeled the graph using Google Slides in Round II.

Received July 2022; revised January 2023; accepted March 2023