"Mirror, Mirror, on the Wall" - Promoting Self-Regulated Learning using Affective States Recognition via Facial Movements

Si Chen
sic3@illinois.edu
School of Information Sciences, University of Illinois at Urbana-Champaign
USA

Yixin Liu
yl118@illinois.edu
School of Information Sciences, University of Illinois at Urbana-Champaign
USA

Risheng Lu
risheng2@illinois.edu
College of Liberal Arts and Sciences, University of Illinois at Urbana-Champaign
USA

Yi-Chieh Lee
ejli.uiuc@gmail.com
Department of Computer Science, University of Illinois at Urbana-Champaign
USA

Yuqian Zhou
yuqian2@illinois.edu
Department of Electrical and Computer Engineering, University of Illinois at Urbana-Champaign
USA

Yun Huang
yunhuang@illinois.edu
School of Information Sciences, University of Illinois at Urbana-Champaign
USA

ABSTRACT

Prior research suggests that affective states of self-regulated learning can be used to improve learners' cognitive processes and their learning outcomes. However, little research explored the effect of using facial movements to detect learners' affective states on self-regulated learning. In this work, we designed, implemented, and evaluated *Mirror*: a self-regulated learning tool that applies facial expression recognition to support learners' reflections in video-based learning. We conducted two studies to identify user needs (with 12 participants) and to evaluate the tool (with 16 participants). The results show that, after watching a video, participants benefited from using *Mirror* through different reflection processes, e.g., gaining a deeper understanding of their learning experiences through *self-observation* and attributing causes for their learning affects through *self-judgment*. Meanwhile, we also identified several ethical concerns, e.g., users' agency of handling the uncertainty of AI, reactivity towards outcome-based AI, over-reliance on "positive" AI results, and fairness of AI informed decision-making.

CCS CONCEPTS

• **Human-centered computing → Laboratory experiments**.

KEYWORDS

Video-based Learning, Emotion, Affective Computing, Mixed Methods

DIS, , June 13 - 17, 2022, Virtual Event, Australia

ACM ISBN 978-1-4503-9358-4/22/06…$15.00
https://doi.org/10.1145/3532106.3533500

ACM Reference Format:
Si Chen, Yixin Liu, Risheng Lu, Yi-Chieh Lee, Yuqian Zhou, and Yun Huang. 2022. "Mirror, Mirror, on the Wall" - Promoting Self-Regulated Learning using Affective States Recognition via Facial Movements. In *DIS '22, June 13 - 17, 2022, Virtual Event, Australia.* ACM, New York, NY, USA, 15 pages. https://doi.org/10.1145/3532106.3533500

1 INTRODUCTION

Self-regulated learning refers to the process that students activate and sustain cognitive, metacognitive, behavioral, motivational, emotional/affective aspects of learning towards their own goals [60, 66, 77]. Learners' affects (also noted as 'learning emotions'), such as confusion and curiosity, are found to be impactful on their learning outcomes [20, 61]. Thus, tools have been developed to track learners' interactions with educational technologies and then provide learners' information and insights into their learning process [55, 83]. For example, visualizations of learners' data (such as screen-recording, concurrent think-aloud, eye tracking, interaction log, facial expression, notes) could be used to support emotion regulation during self-regulated learning [4]. These tools can also encourage learners to adapt their learning behavior, e.g., applying time management strategies [43, 55, 68], and to cultivate learners' meta-cognition skills, e.g., improving awareness of one's own thought processes and gaining a deeper understanding of the patterns behind them [9, 28, 55, 59].

Recent works show that facial expression recognition technologies can be used to detect learners' affective states (also noted as learners' affects) and subsequently can be leveraged to promote learners' reflections [4, 65]. Facial expression recognition is a promising field of computer vision and artificial intelligence (AI) [53]. It becomes increasingly applied for automatic affects detection in online learning [31, 65]. For example, a dashboard is created for teachers to be aware of students' emotions in online learning [31]. A recent study also suggests that students become more aware of their facial expressions in video-based online learning, implying the potential of detecting facial movements for learners to improve their online learning [17].

However, little is known about the effect of using facial expression recognition to detect learners' affects on self-regulated learning. Unlike the common user behavioral data, e.g., frequency of accessing learning documents, time spent on different materials, or grades evaluating learners' mastery of specific subjects or skills [9], learners' affects detected by AI-based solutions often come with errors. Because learners' perceptions of the accuracy and fairness of AI impact the perceived usefulness of AI-based intelligent learning systems [40, 47], it is important to explore how imperfect recognition of affective states impacts learners' use of the AI-based system. Additionally, even though ethical concerns, e.g. privacy and surveillance concerns of learners, are essential when designing learning analytic systems, few studies directly addressed these issues in system evaluation [88].

To fill the void, we designed, implemented, and evaluated a self-regulated learning tool that detects learners' affective state in video-based learning. We call the tool ***Mirror***, hoping that learners could be able to "see" their own learning affects in video-based learning. In this research, we leveraged the use of facial expression recognition [5, 31, 50] to inform learners of their affects. To inform the design, we first conducted a need-finding study with 12 participants to identify key system features and interaction ideas from learners. We then implemented the identified features and conducted a user study with 16 participants to explore how learners would perceive and use *Mirror* for reflections in video-based learning.

Our work makes the following contributions. First, we propose a novel tool for self-regulated learning–*Mirror*, which offers a novel mechanism of detecting and presenting learning affects of video-based learning and supports learners to conduct different reflections on their thought processes via facial movements. Second, we conducted a two-phase study, yielding empirical evidence on how learners conducted multiple reflection processes for video-based learning, e.g., identifying patterns in their learning affects, evaluating their experiences, and attributing causes leading to their learning affects. Meanwhile, we also found that the type of affects (e.g., positive or negative) promoted different reflection processes. Third, we found that when using the AI-based tool, participants had several ethical concerns, including: users' agency of handling uncertainty of AI, reactivity towards outcome-based AI, over-reliance on "positive" AI results, and fairness of AI informed decision-making. Last, our work brings insightful implications for designing AI agents to improve consciousness of reflections in self-regulated learning by mitigating biases and eliciting guiding questions.

2 RELATED WORK

2.1 Tools Supporting Self-Regulated Learning

Self-regulated learning is described as a multifaceted concept that includes monitoring of cognitive, metacognitive, behavioral, motivational, and emotional/affective aspects of learning[60]. It is an iterative and sequential process that includes multiple phases of reflections. For example, the cyclical phase model for self-regulated learning posits a Forethought, Performance, and Reflection phase [93]. Specifically, the Forethought phase involves behavior before learning, such as setting goals; in the Performance phase, learners execute *self-control* processes and *self-observation* processes that provide internal feedback; in the Reflection phase, learners self-evaluate progress informed by *self-observation* and conduct *self-judgment*; further on, learners express emotional responses to the judgment, which serves as input for the next iteration of self-regulated learning.

Learning analytic tools, e.g. [33, 66], allow learners to monitor their learning processes and to facilitate reflection in 'self-regulated learning'. It is found that lifelong skills such as "meta-cognition" (the monitoring and control of thought, as"thinking about thinking" [59]) could be gained through the use of learning analytic tools [43, 55]. It helps learners achieve their learning goals through better awareness (observing their data), reflection (asking themselves questions), sensemaking on learners' own data (e.g., answering questions asked), and intention towards to behavior changes (e.g. such as improvement in time management strategies) [43, 55, 68, 83]. Self-regulated learning frameworks, such as the cyclical phase model for self-regulated learning [93], are often applied for scaffolding learners' reflection that is supported by learning analytics tools [56, 67].

Many learner-facing learning analytic tools only track and visualize learners' behavior/performance that can be explicitly logged by the computer systems. Only a few of them focus on facilitating learners' reflection on affective states that are essential for self-regulated learning. For example, according to a research review [9], 75% of the learning analytics tools studied "resource use" (e.g., the number of times a resource was accessed); 37% included "assessment" (e.g., learners' mastery of certain skills or subjects as measured by assessment instruments); and 30% addressed "Time spent" (e.g., the amount of time spent accessing resources) [9, 31].

2.2 Recognizing Learners' Affective States

Recently, scholars suggest that online learning systems should better understand learners' affective states [65], e.g., by leveraging learners' facial expression recognition in learning analytic dashboards [4]. For example, *Emodash* is a teacher-facing learning analytic dashboard, which tracks learners' facial expressions to recognize basic emotions, including: anger, contempt, disgust, fear, happiness, neutrality, sadness, and surprise based on Ekman et al.'s classifications [29]. The distribution of learners' positive and negative emotions are displayed for teachers and tutors to increase their retrospective awareness of learners' emotions in a video-conferencing learning environment [31]. Unlike *Emodash* that focuses on teachers' awareness of students without evaluating learners' use of the tool, our study targets learners' own awareness of their emotions during self-regulated learning using videos.

Although experienced teachers are adept at recognizing the learners' affects through facial movement observations, learners themselves cannot accurately track their emotions, which further impacts their learning engagement and outcome [4]. Automatically recognizing learners' cognitive-emotive states (affective states) by analyzing their facial movements can be supported by intelligent learning systems [50] built on advanced computer vision and AI systems [53]. These systems are designed to present the learners' affective states in an interpretable way and to help understand learners' cognitive models and interactions. For example, some works found that neutral and positively-valenced (e.g., happy) represented the majority of learners' emotional states experienced with *MetaTutor* [36]; and facial expression data together with eye-tracking and note-taking can potentially model learners' cognitive processes [3].

Noteworthy, previous research found that the predictions of facial expression recognition algorithms do not infer deterministic emotions, and the results are not solely determined by facial movements, but also related to other influential factors like head poses and social context information, etc [6]. This means affects recognized via facial movements may not perfectly capture learners' affects and match with the their own interpretations. Therefore, it is important to understand how learners perceive the accuracy and usefulness of AI-based learning systems [40].

2.3 Ethical Concerns with Learning Analytics

Ethical concerns have been widely discussed in the field of AI research for education. For example, data ownership, users' expectation of privacy, transparency and intelligibility of decisions informed by AI impact are critical factors to consider when applying AI-based technologies for education [88]. On the one hand, AI offers the hope of increasing personalization in education and complementing the work of (human) teachers without dispensing with them; on the other hand, it is accompanied by risks of becoming less social and raising privacy concerns [70]. By tracking students' behaviors, rich data could be collected (e.g. facial movement [9]); however, people questioned the surveillance on teachers and students [15, 70]. There are also on-going critiques on learning analytic research being "solutionism" , in which researchers treat technologies as solutions by assuming all innovations are necessarily good [79, 89]. Researchers and practitioners in education community should reflect on the ways in which AI might could people's lives and to embrace their responsibilities to enhance its benefits while mitigating the potential harms [12, 14]. Technologies are rarely good or bad in themselves and what matters is how they are used [14, 38, 70]. To improve practices of deploying ethical education technologies, human-centered design should be applied. It give stakeholders the genuine agency in shaping digital tools and providing their own insights, such that AI-based learning tools can proactively address major ethical concerns [38, 70]. Additionally, designers, data scientists and other stakeholders in learning analytics should reflect on power relations between them and those who traditionally have been the subjects of their research, e.g. asking what value system are encoded in learning analytics [89].

Given the above literature, we aim to address the following research questions:

RQ1: *How do learners use affective states recognized by facial movements for reflection in self-regulated learning?*

RQ2: *What are learners' perceptions of using facial movements recognition for reflection in self-regulated learning?*

RQ3: *What are ethical considerations of using affective states recognition for reflection in self-regulated learning?*

3 *MIRROR* DESIGN

In this study, *Mirror* is designed for video-based learning. This is because video-based learning is self-paced and widely used for online education. It is important to provide support for improving self-regulated learning strategies, which often improve learner engagement and learning outcomes [45]. Many learning analytic tools are designed for video-based learning [21, 26, 31], and recent works suggest to leverage learners' affective states for improving self-regulated learning online [17]. Our work fills the void for designing learning analytic tools with affective states recognition via facial movements.

Affective states are typically represented in the form of discrete emotion categories [30] or dimensional models incorporating continuous arousal-valence values [73]. James Russell *et al.* proposed the circumplex model [73] that distributes human emotions in a two-dimensional arousal-valence circular space. Notice that arousal measures the intensity of an excitement ranging from -1 (indicating calm) to 1 (indicating excited), while Valence decides on the pleasantness from -1 (very negative) to 1 (very positive) [7]. The 2D space formed by these two values explicitly represents the affective status of humans [73]. Compared with discrete emotion labels, continuous measurement shows the dynamic and gradual changes of emotions to analyze emotional states' intensity and transitional behaviors. Therefore, the circumplex model has been widely used for measuring and evaluating affective states [71]. Recent AI technologies enable the prediction of arousal-valence from facial movements [24, 25, 49, 63].

We then applied a state-of-the-art arousal-valence regression algorithm that returns a series time-based arousal-valence pairs [24]. We chose this technique because it outperformed several tools in the FG2020 challenge on Affective Behavior Analysis in-the-wild (ABAW) [49]. The tool consisted of a face detection and an arousal-valence regression module. Specifically, the face detection module was implemented using MTCNN [92]. Given a sequence of consecutive video frames, our system detected and cropped the faces, and analyzed the facial images in the regression module. Finally, the system returned two per-frame sequences of arousal and valence. We utilized the algorithm to process learners' face image sequence and post-process the AI-recognized results in the form of arousal and valence to reflect learners' affective states.

Additionally, previous research suggests that measuring emotions as solely objective measurements, such as facial movement, is insufficient [80]. Instead of only sensing and transmitting emotion, systems should support human users in understanding, interpreting, and experiencing emotion in its full complexity, and avoid ambiguity [10]. Our system also had participants provide self-reported affective states through text alongside the AI's predictions to mitigate the issues.

3.1 Need-finding Study

To understand what affective states might be recognized by the AI tool and how to visualize the recognized affective states, we recruited 12 participants (4 Female, 8 Male) to conduct a pilot study. Three of them are undergraduate students and nine of them are graduate student at a US public school. In the study (40-60 minutes each), participants were asked to record their faces while watching two videos: a self-selected video using the camera on their laptops (10-15 minutes), and an assigned learning video that intended to elicit viewers' different affective states (15 minutes). We compensated each participant 10 USD per hour and conducted under approval by our university's Institutional Review Board. Below, we present details of each session. The assigned video is about introducing Augmented Reality (AR), which was the same one we used for tool evaluations. Participants selected the video based on the topic they are interested in, but there could be still content in the video that is boring. Participants selected videos such as cooking tutorials, public speeches, and class material etc. Participants were asked to watch the whole video, though they could pause/resume watching the video (when a student pauses/starts the video, the facial movement detection is paused/resumed simultaneously). The participants were asked to choose the videos' timestamps and share their perceived affective states for both videos with researcher. The reason for asked participants to recall their affective states prior to showing AI results to avoid their over-reliance on AI results. Similar approach are also used in other learning systems that prompt user to self-assess before seeing AI-results, e.g. guitar learning [85].

Then the researchers showed participants two visualization approaches using the recognized affective states: 1) *Scattered Plot* and 2) *Line Chart*, as shown in Figure. 1. The rationale of the two visualization approaches is as follows.

The *Scattered Plot* was designed to show the distribution for each pair of arousal-valence values in a 2D space. We leveraged the continuous rainbow legend for visualizing arousal-valence pairs by referring to [39, 52]. Specifically, we used cool colors for negative results and warm colors for positive results. Following the previous paper, we remapped the pair values in the arousal-valence 2D space into polar space represented by intensity and angle. Finally, we plotted the data points of intensity in a 1D graph using color in the HSL color wheel indicated by the angle. Showing distribution using a pie chart is widely used in dashboard designs to provide an overview on learning performances [31, 55].

The *Line Chart* was designed to show the change of arousal-valence values in two separate lines throughout video watching activity. Plotting 'Arousal' and 'Valence' in two separate lines is often used to visualize facial recognition results returned using the arousal-valence model [24, 25]. Using line charts to visualize changes is often applied in dashboard designs to facilitate reflection on one's own learning process, e.g. [31, 55].

The participants were asked to share which *visualization* they preferred for reflections. All participants shared that they thought the line chart was more helpful, allowing them to identify the changes of emotions over the video display. Therefore, we chose to visualize the AI results with using a line chart.

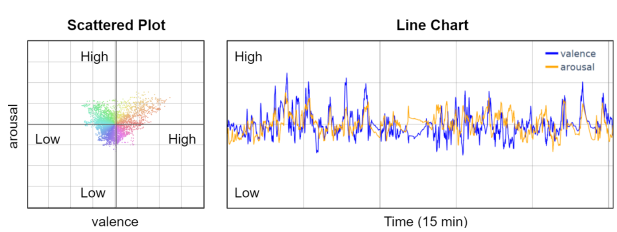

Figure 1: Two visualization options explored in the need-finding study: *a scattered plot* (left) and *a line chart* (right). Participants were able to see details of the data points by zooming in on a particular area.

The participants were also asked how they would like to use the visualization to support their reflection of learning. Eight out of 12 participants found it hard to translate the two measurements together (Valence line and Arousal line) and interpret two values at a specific time point in the line chart shown in Figure. 1 (Right). These participants suggested that the system should merge the two lines into one line and return adjectives that describe affective states. Meanwhile, they suggested 'arousal' should be revised to more context related terms, such as 'Engagement Level'; found 'valence' related terminology 'negative/positive' which is directly used from previous paper [7] intuitive and easy to follow. Such feedback indicates that the 2D space for arousal-valence should be represented in one 1D line-chart and intuitively show 'engagement' changes between different learning stages. Therefore, the final visualization of the recognized affective states is designed as area 2 of Figure. 2.

More specifically, to better illustrate the two values in a more concise and intuitive way, we remapped the pair values at each timepoint using function used in *Design 1 Scattered Plot* and used different colors to represent recognized learning states at each timepoint. One of the examples of a time-based curve is shown in Figure. 2. The height of the data points demonstrated the intensity of the specific states, and the color showed the recognized affective states, which could be more easily understood by the users. According to our participants' feedback, 'arousal' intensity changes over time was considered to be more informative and relevant to in-class performances then 'valence'. Therefore, 'arousal' was finally selected as the y-value.

3.2 Supporting Learners' Reflection with the Affective States

We synthesized participants' feedback and identified several features to support learners' reflection after watching videos. A desktop view of *Mirror* interface is illustrated in Figure. 2. The tool was presented to learners after they finish watching the videos.

The design rationale of *Mirror* is as follows. First, to help users make sense of the data, changes between different learning stages and corresponding legends of arousal-valence 2D space are provided on the dashboard, as shown in Figure. 2. Second, all participants found it is essential for them to understand the system's performance by mapping AI results with their own perceptions. They explained that, by finding strong matches between the recognized states and their own perceptions at aligned time points, they were more likely to trust AI at time points where they cannot recall their own perceptions. Additionally, a recent study on capturing cognitive, motivational, and emotional learning processes suggests that tracing learning activities with objective data (e.g. facial recognition) and combining those traces with subjective (e.g. self-reported) data may provide new insights on how users learn in online environments and inform online learning system design [65]. Informed by such findings, the participants were given a table to select time-points to fill in their self-reported reflections after finishing video watching but prior to seeing the interface shown in Figure. 2. Once the self-reported affects were collected, they were pinned on a timeline under the affective states graph (as shown in area 3 of Figure. 2) on *Mirror*. To support learners' recalling the

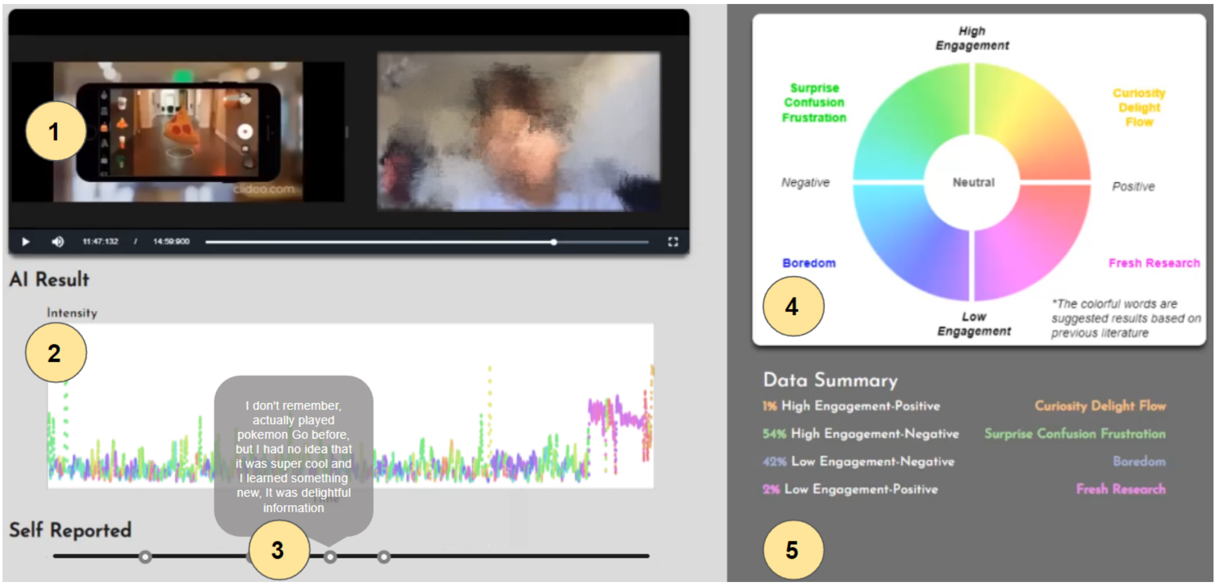

Figure 2: After watching the video, *Mirror* participants were given the above interface. They could navigate the five areas illustrated above: (1) displaying the video recording next to the participant's video selfie when she was watching the video, (2) visualizing time-based AI results of affective states via facial movements, (3) tagging self-reported moments of affective states, (4) providing a legend for Time-based AI Results, and (5) summarizing AI results. The sample data was from P2.

corresponding video content and video selfie, the interface is designed to display the video lecture and video selfie based on the timestamp of the self-reported states or the moments on the AI graph, as shown in area 1 of Figure. 2.

4 METHOD

4.1 Study Design for Evaluating *Mirror* Use

To understand how learners would use our tool for video-based learning, we conducted a three-session user study, as illustrated in Figure. 3. We recruited 16 participants (10 female, 6 male, aged 19-28). Eight of them were undergraduate students and eight of them were graduate student at a US public school. Because the video to be used in the study was about introducing Augmented Reality (AR), we selected participants who had no prior background in AR. We compensated each participant 10 USD per hour. Each study took about 1.5 hours (Figure. 3). The studies were conducted and recorded through the video-conferencing platform Zoom. The study was approved by our university's Institutional Review Board. Below, we present details of each session.

4.1.1 Session 1: Video Learning and Self Reporting. In the first session, researchers asked participants to watch an assigned video about AR concepts and applications as if they were learning online by themselves. To keep participants' engagement in watching the video, we followed a prior study's suggestion [35] by selecting a 15-minute long video lecture from Coursera to inspire the viewers' various affective states in learning. While participants were watching the video, they were asked to open their cameras, and the researcher recorded participants' faces, having their consent in advance. Participants were asked to watch the whole video, though they could pause/resume watching the video. When a student pauses/starts the video, the facial movement detection is paused/resumed simultaneously. Before the video watching started, researchers showed a demo on how the tool works.

After watching one video, the participants were immediately asked to report their affective state for the video. Writing post-learning reflections has been found helpful for learners in terms of actively reviewing learning material and summarising takeaways [1, 19]. Also, the self-reported reflection allows researchers to analyze whether AI's results match users' own perceptions and further form interview questions. Participants were asked to select moments on the video timeline to write their reflections while associating to the video content. To compare their self-reported affective states with the AI results recognized by *Mirror*, participants were given nine states as options without being required to use them, including *confusion, frustration, surprise, curiosity, delight, flow, fresh research, bored, neutral*. The states' options were presented in a Valence-Arousal matrix that has been widely used in previous models to predict affective states in learning [42, 51]. We did not limit participants' words to describe their emotion states because different AI models may cover different adjectives to predict affective states, e.g., misconception, discard, etc. [51, 62].

4.1.2 Session 2: Thinking Aloud while Using Mirror. In the second session, following previous research that studies how users interact with personal data visualization [18], firstly, a researcher demonstrated how to use the interface and then talked about how the researcher synthesized insights from our interface using 'think-aloud.' Secondly, the researcher showed the participants' data in the *Mirror* interface by sharing the researcher's computer screen remotely; then the researcher asked the participants to control the interface remotely. Thirdly, the researcher asked the participants to 'think aloud' as they explored the *Mirror* interface to understand thought processes while exploring the learning analytics and insights they gained from the exploration of the *Mirror* interface. During the think-aloud protocol, the researcher observed how the participants interacted with the interface and occasionally prompted the participants to think aloud and asked clarification questions.

The think-aloud sessions lasted between 10 minutes to 40 minutes and were video recorded using Zoom. The recordings including screen recordings of the participant's interactions with *Mirror*, their speech, and participants' video selfie (facial expression). The participants were allowed to use the *Mirror* interface as long as they needed, to describe their observations sufficiently.

4.1.3 Session 3: Exit Interview. In the third session, inspired by a previous research [18] to investigate users' interactions with personal data visualization, we asked participants to answer questions related to how they perceived the user experience of interacting with *Mirror* and factors that impact such experience. This session lasted between 15 minutes to 40 minutes. The interview sessions were video recorded on Zoom.

4.2 Data Analysis

To address RQ1 about participants' use of *Mirror* with self affective states, we referred to the previous paper [18] to firstly establish properties of what participants said without relying on existing theories (open coding), and proceeded to identify relationships among the codes (axial coding) [18]. This approach allow us to understand how participants reflected on affective states in learning through visual exploration using our tool (Session 2 think-aloud). Two researchers read and coded 30% of the transcripts for each session independently, and their initial inter-rater reliability (observed proportionate agreement) was 79%. They discussed discrepancies and revised and expanded the existing categories until reaching an agreement. They then coded the rest of the data individually. To address RQ2 - how participants perceived the use of *Mirror* , we coded participants' interview feedback by following the same process as coding the think-aloud results. The initial inter-rater reliability was 81%, and the researchers revised and expanded the existing categories until they reached an agreement.

5 FINDINGS

5.1 Learners' Reflections Promoted by Their Affective States (RQ1)

In this section, we present our findings to address the proposed RQ1. During the think-aloud session (as shown in Fig. 3), each participant interacted with *Mirror* and reflected on multiple segments of the video. Three main types of user interactions with *Mirror* involved participants' reflections: 1) relating to corresponding learning content by rewinding the video (area 1 in Fig. 2), which happened 111 times; 2) reviewing Mirror graphs for the recognized affective states (area 2 in Fig. 2), which happened 67 times; and 3) checking the self-reported states (area 3 in Fig. 2) and associating them with the affective states video selfie, which happened 53 times. We observed 231 interactions among 16 participants.

Participants had in total on 212 reflection segments within 231 interactions. Each reflection segment starts with one interaction with *Mirror* and may further incorporate other interactions in the same sentence. For example: P2 clicked a peak in the graph (area 2 in Fig. 2) but forgot the video content at that timestamp, so they rewinded the video to review (area 1 in Fig. 2) and reflect in the same sentence *"let's see this green spike over here (review graph), but I kind of forgot and let's see what happened (rewind video). Yes, it is about using AR for nonprofit and that was something I didn't know before watching the video. "*.

For each reflection segment, their think-aloud input allowed us to identify which specific reflection processes were involved. More specifically, according to the cyclical model on reflections in self-regulated learning by Zimmerman *et al.*, learners' reflections with *Mirror* were coded in alignment with the cyclical phase model of SRL [93]. More specifically, participants' reflections happened at two processes: 1) ***Self-Observation***–"tracking of specific aspects of their own performances, the conditions that surround it, and the effects it produces"[93]; and 2) ***Self-Judgement***–"self-evaluating one's performance and attributing casual significance to the results"[93].

5.1.1 Mirror Supporting Multiple Reflection Processes. When participants interacted with *Mirror*, ***Self-Observation*** happened 189 times (89% of the 212 reflection segments). Within *Self-Observation*, we further identified two types of reflection subprocesses according to prior literature [18]:

- *Identify details*: during this subprocess, participants tried to reflect on details of an individual learning segment that triggered the affective state. It happened in 124 (65%) of the total reflection segments (on average eight times per participant).
- *Compare states*: during this subprocess, participants compared states at different timestamps along the video display to draw insights on their learning experience. It only happened in 26 (14%) of total interactions (on average twice per participant).

Besides self-observation, participants further proceeded with ***Self-Judgement*** reflections, which happened 144 times (68% of the 212 *Mirror* reflection segments). Within *Self-Judgement* interactions, three types of reflections are further identified.

- *Recall External Context*: Participants recalled external context to explain what they observed, such as what caused them to feel bored, which happened in 139 (66%) reflection segments; almost all *Self-Judgement* reflection segments had such reflections (on average nine times per participant).
- *Self-Evaluation*: Participants made evaluations on whether or not their behavior matches their goal. It only happened in eight reflection segments (on average 0.5 reflection segments per participant).
- *Attribution*: Participants reflected on what caused the results of *Self-Evaluation* and then made further plans. It only occurred in seven reflection segments (on average 0.4 reflection segments per participant).

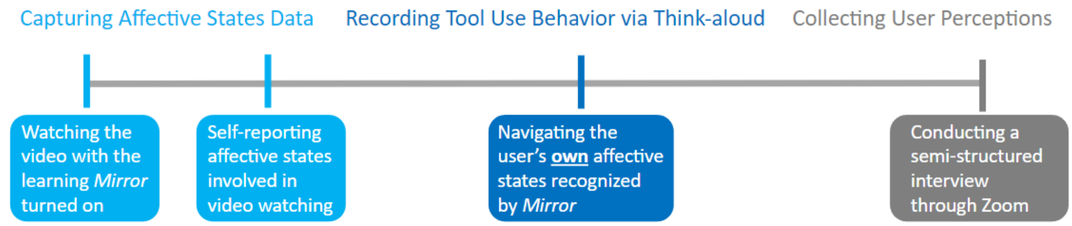

Figure 3: Three sessions were included in the study: 1) collecting learners' affective states that were either automatically recognized by the tool or self-reported by the participants; and 2) recording tool-use behaviors via think-aloud while users used the *Mirror* interface for reflection after watching one video; and 3) collecting participants' perceptions of their tool-use through interviews.

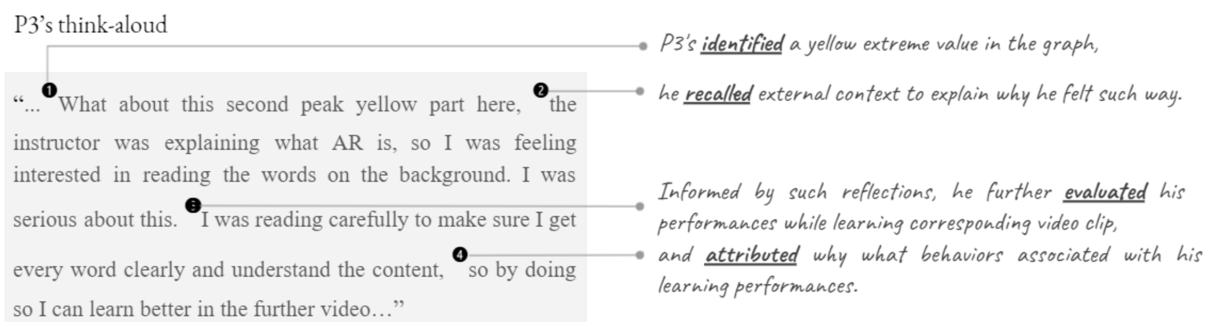

Figure 4: Sample data and analysis results: P3's original think-aloud transcripts (Left) and researchers' narratives for explaining the participant's reflection processes (Right).

Note that one *Mirror* reflection segments could have involved several reflection processes. For example, P3's think-aloud provided an example (as shown in Figure. 4).

We then conducted a correlation test to examine the relationship between *Mirror* feature use and reflections processes. The result showed that more types of interactions with *Mirror* were associated with more subprocesses of reflections (S=174.7, p< .001). Namely, participants who used all the *Mirror* features (e.g., reviewing the graphs, checking self-reports, and rewinding videos to relate their emotions) conducted more diverse reflection subprocesses.

5.1.2 Reflection Processes Varying by Recognized Affective Valence.

Among the 212 reflection segments, 37% were related to timestamps where participants showed negative valence (mean=4.9 per participant, SD=3.4), displayed in cool color as the legend shows in Fig.2 (area 4); 23% were related to timestamps where participants showed positive valence, displayed in warm color in the visualization (mean=3 per participant, SD=2.4); and 40% were recognized as Mixed learning timestamps, where warm and cold colors were mixed (mean=5.4 per participant, SD=4.1). The duration of each interaction is input by participant think-aloud data.

It appears that affective valence leads to different reflection processes. In particular, our correlation test result showed that more types of *Self-Observation* subprocesses were associated with more interactions with *Mirror* on negative learning timestamps (S=254.4, p < .05), not with positive learning timestamps. On the contrary, more types of *Self-Judgement* subprocesses were associated with more interactions with *Mirror* on positive learning timestamps (S=381.5, p < .05), not with negative learning timestamps.

During the think-aloud process, participants considered the 'valence' value of high importance for reflection and easy to visually capture. All participants referred to negative 'valence' as parts worth to further conduct reflection than positive 'valence,' as they were concerned that they *"might be missing something"* or *"they might have not understood parts of the video."* Additionally, our correlation test showed that more interactions on negative learning affects were associated with more types of interactions with *Mirror* (S=293.1, p < .05), while there was no correlation between positive learning affects and interactions with *Mirror*.

Summary When *Mirror* visualized participants' own affective states that were automatically recognized via their facial movements, the proposed visualizations and interaction features successfully promoted participants' multiple reflections, e.g., two subprocesses of *Self-Observation* and three subprocesses of *Self-Judgement*, on their video-based learning. For some participants, affective valence (negative and positive) had an impact on the type of reflections.

5.2 Technical and Psychological Factors Impacting User Experience (RQ2)

In this section, we qualitatively analyzed think-aloud data and interview feedback to develop further understand of participants' user experience. Overall, participants found the tool to be helpful for supporting reflection process and self-regulated learning experience. Then, we zoom in on factors that negatively impact user experience.

5.2.1 Supporting Reflections though Affect Patterns. Participants often started with *Self-Observation* triggered by the *Mirror* visualizations, then proceeded with *Self-Judgement*. This sequential relationship was aligned with the cyclical model of self-regulated learning proposed in [93]. While conducting *Self-Judgement* using *Mirror*, participants may proceed with new *Self-Observation* reflections.

For example, Figure. 6 shows a case provided by P12, when a sub-process (i.e., *identify details*) of *Self-Observation* was conducted after a sub-process (i.e., *recall external context*) of *Self-Judgement*. Peaks as arousal pattern attracts learner's attention to proceed to the next reflection segment. Interview results showed that sharp changes of arousal, both warm-colored peaks and cool-colored peaks, indicate fully dedicated to the learning process. In other words, high engagement negative emotions are also considered valuable and unnecessary to avoid in learning.

We further examined whether users gained insights from valence patterns. Nine of the 16 participants compared the valence changes on their graphs over time. The nine participants that compared time-wise changes on average had 29 reflection segments; while the other seven participants only had 13 reflection segments. Comparing the changes of their states from negative to positive or vice versa allowed some participants to reflect on the causes of these changes (conducting *attribution* reflection as *Self-Judgement*). According to prior literature [66], attributions help people plan for actions for making self-regulated learning more effective. For example, when P10 compared the facial expression results, she was able to realize a phone alarm disturbed her learning through the first half of the video watching, as our narratives explained in Figure. 5. On the contrary, in her self-reported states, she logged many positive affective states for that period, as shown in Figure. 5). Interview results showed that at the very beginning of the think-aloud session, participants first looked for time-wise changes in graph before relating with any video content or checking self-reports. Further on, as participants zoomed into specific time points, they paid less attention to the overall time-wise changes.

5.2.2 Improving Meta-cognition Skills in Self-Regulated Learning. In brief, interview results show that all participants found the tool to be helpful for improving *Cognitive Self-consciousness* , defined as "the tendency to focus attention on thought processes" [87], improving *Cognitive Confidence*, defined as "confidence in attention and memory" [87], and improving *Positive Belief*, defined as "concerning the usefulness of rumination" [87]. We organize our findings below in two layers and explain how different features contribute to the self-regulated learning: regulation of processing modes and regulation of the learning process in alignment with [11].

Participants found that tracking affective states can help them release cognitive resources stuck with difficult learning material. Long periods of negative valence together with high arousal were mentioned by four participants that they may have spent too long on a preceding concept and overlooked upcoming concepts. By being more aware of such behaviors during their self-regulated learning and knowing they were able to trace back later on, they imagined themselves to be learning efficiently and avoided rumination

Participants found that the tool nudged them to examine their emotions in learning activities and to be more thoughtful of their behaviors, e.g. when to take note, when to review. Additionally, the AI-suggestion in alignment with their self-reported result enforced them to be cognitively engaged in the reflection activity. They suggested that by periodically using the tool, they would commonly reflect on learners experience in-class and after-class.

5.2.3 Factors Negatively Impacting Learners' Reflection. First, during interview and think-aloud session, four participants shared that it was challenging to use *Mirror* for *Self-Observation* reflections, e.g., conducting time-wise comparison, because it was hard to visually identify patterns of the recognized affects. One major reason was *"information overloading for general learners to make use of the information."* They suggested that further design should *"smoothen the line"* (P2) and *"automatically mark-up changing points"* (P5, P6, P7). In brief, they preferred less affect states to be visualized on the video timeline to reduce the cognitive load and subsequently to improve the reflection efficiency.

Second, misalignment between continuous AI results and discrete self-reports interrupted learners' *Self-Observation* reflections. Nine out of 16 participants located 16 *"engagement peak"* and 12 *"positive/negative changing timepoint"* that lacked self-reports to handling ambiguous situations in reflections. Although participants rewinded video selfies and checked self-reported reflections (for example P11 in Figure. 7), some of our participants (P4, P5, P9, P11) still could not resolve mismatches and had confusion. They tangled repeatedly by replaying video segments multiple times to attribute whether the AI was correct and paid less attention to either the video content itself or their own learning behavior, and they were not able to proceed to *Self-Judgement*. P3 recommended that *"when video learning content gets longer, and the video content gets complicated, they will have less confidence in their own perceptions of affective states at both timepoints with/o self-reported reflection."*

Summary Visualizing changes in valence and arousal patterns had positive effects on participants' reflections, e.g., raising awareness of important concepts that might be overlooked, and driving cyclical reflections with new discoveries about their own learning experience. The alignment and misalignment between self-reported affects and AI results encouraged learners to examine their feelings and were perceived to encourage meta-cognition processes. Information overloading, as a result of large amount of data accompanied with the small granularity, was found to negatively impact some participants' perception of *Mirror* use.

5.3 Ethical Concerns of Using *Mirror* (RQ3)

In this section, we qualitatively analyzed the think-aloud and interview data to understand ethical considerations of using *Mirror*. The analysis below does not intend to compare different presentations themselves, instead it focuses on different interactivity enabled by the two presentations in *Mirror* supported and hindered ethical reflections.

5.3.1 Learners' Agency When Handling Uncertainty of AI. In *Mirror* interface (area 5 in Figure. 2), the tool showed the percentage of

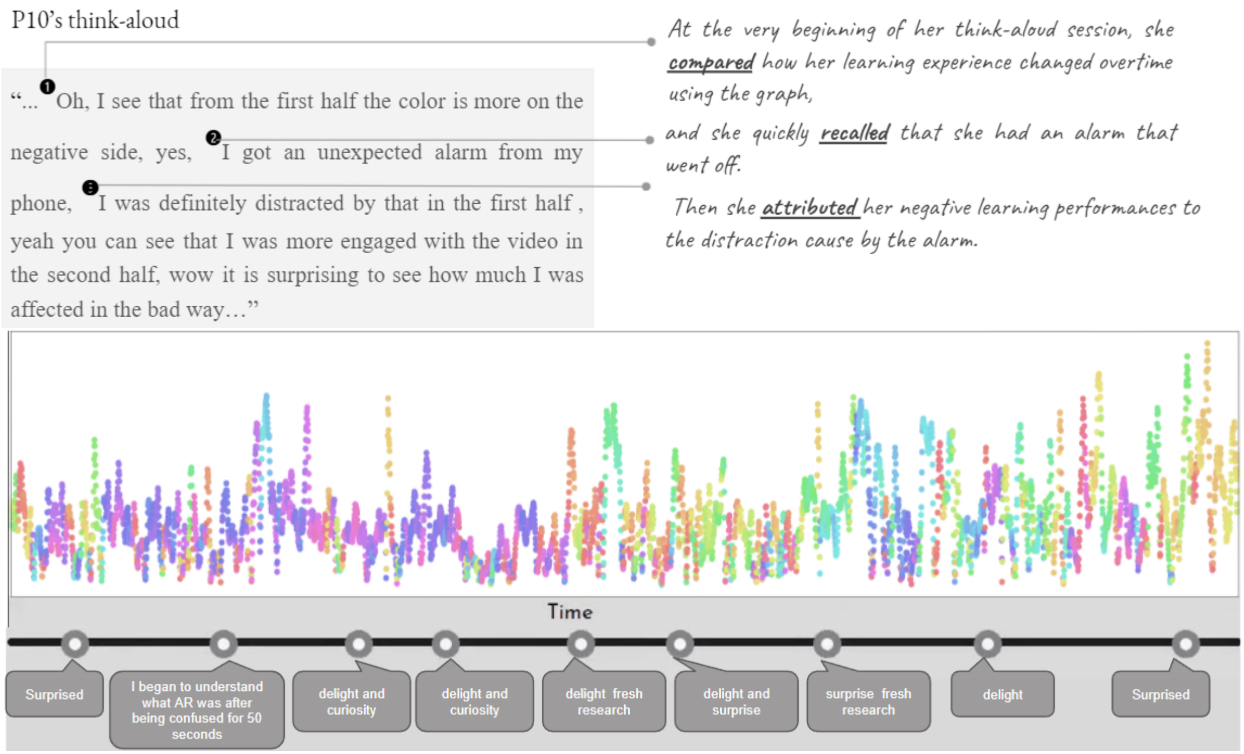

Figure 5: When P10 found that the graph had a changing point in the middle of the video, she quickly recalled and attributed that a phone alarm disturbed her when watching the first half of the video. This was a case of *Self-Judgement* reflection, i.e., by conducting a time-wise comparison between affects on the video timeline, participants could identify major changes of their learning affects and were able to reason about the change.

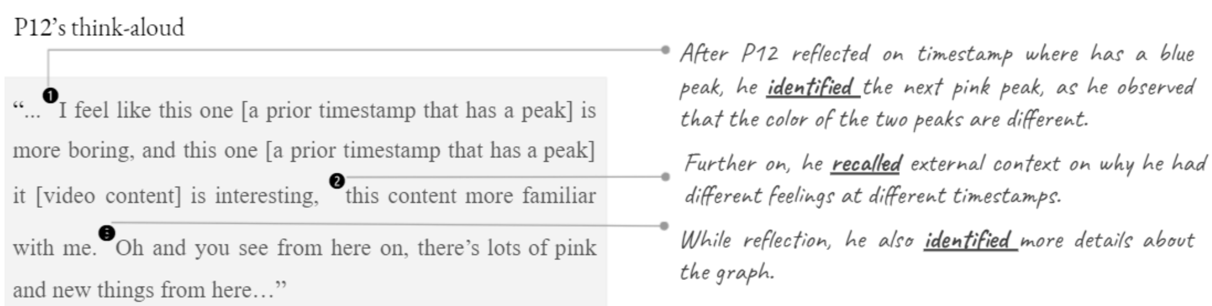

Figure 6: Sample data and analysis results: P12's original think-aloud transcripts (Left) and researchers' narratives for explaining the participant's reflection processes (Right).

frames falling in four quadrants (High Engagement -Positive, High Engagement-Negative, Low Engagement-Positive, Low Engagement-Negative) similar to previous research which often calculate the four percentages for researchers to analyze peoples' affective states [74]. However, users did not consider using four percentages summarizing affective state shown on the interface as helpful in the reflection process. The time-series data showing more granular time series data results in area 2 in Fig. 2 were perceived to better support the reflection than directly giving data summary.

Participants liked the time-series granular visualization because they could handle the perceived errors of AI by themselves. To identify the errors, participants consulted with self-reported data and video selfies when they felt unsure about the detected affects. According to their think-aloud data, 15 out of 16 participants checked the self-reported states and associated them with video selfies for a total of 53 times, and 10 out of 16 participants checked video selfies for a total of 30 times. For example, P4 explained how interacting with the AI results allowed him to use the AI results for reflections better: *"I didn't know how well this tool worked for me even though I*

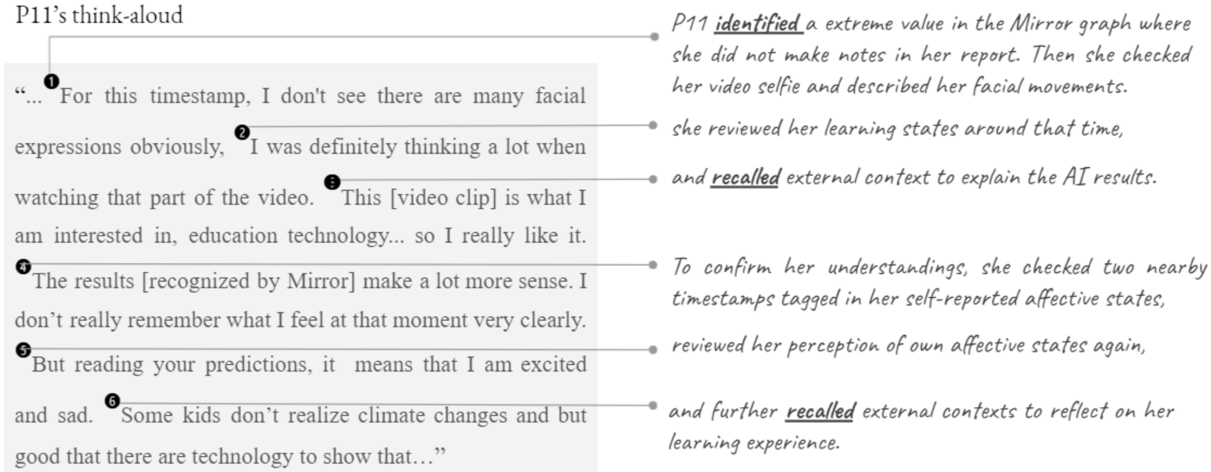

Figure 7: Sample data and analysis results: P11's think-aloud transcripts (Left) and researchers' narratives for explaining the participant's reflection processes (Right).

have taken some AI classes. At first, I was quite skeptical; I could not take the summary as it was, you know; it was like the system made its judgement by itself. But as I explored the visualization, I felt much more inviting when I could quickly re-play the video clips and decide on my own when I felt the AI was not doing well. "

5.3.2 Learners' Reactivity toward Outcome-based AI. Another reason is that the granular visualization supported reflection procedure by nudging proactive reflection actions, compared to AI summaries that were primarily outcome-based. Specifically, the granular visualization showed subtle changes which intrigued the participants to explore the affects and to re-play the corresponding video clips immediately if needed. In comparison, the AI summary only showed the outcome and did not support alongside actions. Moreover, not supporting the procedurality nature of reflection were perceived to further produce reactivity behaviors to AI tracking – not just ignorance of AI, but the annoyance of AI. For example, P1 recalled an annoying experience with AI tracking technology that did encourage reflection behaviors and only highlighted AI as the outcome: *"In my last internship, the company tracked my mouse clicks when I worked from home... I hated it when they sent me notifications on better or worse clicks, then it made compared to previous days...It is not helpful to sentence me what I have done and not telling me what can be improved. I was like 'so what?' ... I feel it intended to make me feel bad about myself, and not willing to help me."*

5.3.3 Over-reliance on "Positive" AI Results. Positive segments could potentially bias participants' reflections. Participants shared that when they saw "positive" AI results, even when they self-reported negative emotions, they tended to agree with the system, and tried to justify their agreement. This tendency could result in *"subconsciously agreeing with the system."* P2 reflected on such behavior in the interview: *"Before you asked me that question on what I was feeling at that self-reported boredom point, I just took the 'confirmation' for granted. I was aware that I started to search for confirmations from the AI, to let them tell me I am doing a fantastic job. "* Another participant also subconsciously reflected on her positive/high engagement time points: *"I was looking in the sequence of High-Engagement and Positive, then High-Engagement and Negative, OR Low-Engagement and Positive. I did not pay any attention to when the AI told me I was Low-Engagement and Negative."* When the researcher asked her why she did not mention reviewing low-engagement and negative emotion, she explained, *"there are mismatches for all emotions, but I am more likely to agree with AI when there is a mismatch for those listed towards the front, they are good results. "* The utterance suggesting that participants might be biased by positive confirmation bias, impacting the learners' use of *Mirror*.

5.3.4 Fairness of AI Informed Decision-Making. In our context, tracking own affective states via facial movements is to support reflection in self-regulated learning. However, six participants expressed concerns towards potential learner-learner and learner-crowd comparison, either picked up by algorithms or by other stakeholders that had access to such data. Learner-learner and learner-crowd have been widely adopted in a large percentage of current learning analytic tools [60]. Participants found that expressiveness through facial movements may differ from one individual to another; as shown in Figure. 8, some expressed 'Valence' more while some expressed 'Arousal' more. Such within-subject differences would further impact the fairness of others' decisions made based on AI results. For example, P16 said *"It can definitely help with individual stuff. But if teachers use that to grade students, or like job interviews are using that, it is not that fair, you know... people expressing differently."*.

Further on, participants not knowing *"others stakeholder's familiarity with the imperfectness of AI results"* is a factor that explains some participants' concerns towards others' access to AI-based results (*" ... I know to what extend I should trust the results as I get to explore it, but who knows whether others also know AI is imperfect* "(P4)). One participant (P1) specifically mentioned that she would only want to disclose selected positive learning affects with others, including peers, teachers, intelligent systems that need such data as input. She thought positive emotions were more likely to indicate

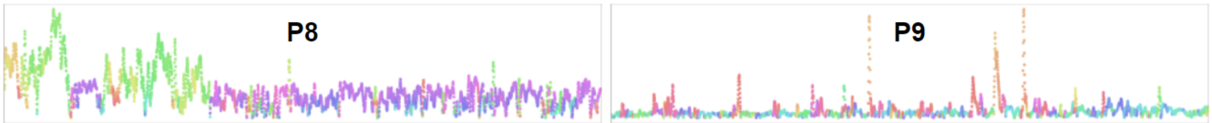

Figure 8: Participants found that expressiveness through facial movements might differ between individuals. Some expressed both 'Arousal' and 'Valence', which is shown as both color changes and intensity change (P8 on the left); however, some only expressed 'Arousal', shown mostly as intensity change (P9 on the right),

that she was doing well and were less likely to lead to harmful interventions: *"... negative emotions are more likely to be used as things against me, it's like a flaw that can be quantitatively compared with someone else. And when people talk about the negative, they by default block out the neutral states..."*.

Summary Our findings show that participants had more agency when AI allowed them to check potential errors by themselves (e.g., consulting with video selfies to evaluate their affects). They wanted to use the affects to better reflect their learning experiences at different times of the video, instead of being provided by a summary of their affects that could be misused for evaluating their learning outcomes. They also shared fairness concerns of using their affects to inform decision-making by other stakeholders.

6 DISCUSSIONS - DESIGN IMPLICATIONS

All the findings show that facial recognition AI can be applied to promote participants' reflections on video-based learning by visualizing learners' affective states, and their reflections could be triggered by cues such as arousal patterns and valence changes. Meanwhile, through a mixed-method approach, we also found several technological and psychological factors (RQ2 and RQ3) had a negative impact on participants' use of *Mirror*. In the remainder of this section, we provide design implications for designing agents to help learners conduct conscious reflections in self-regulated learning. We also discuss future designs of learning analytic tools that are ethical and affective-aware.

6.1 Designing for Consciousness of Self-Regulated Learning

Prior literature found that reflections of self-regulated learning can happen at three phases: *forethought* (e.g., analyzing tasks, setting goals, and planning), *performance* (executing tasks, monitoring progress, keeping cognitively engaged) and *self-reflection* (assessing performance, and making attributions about their success or failure) [93]. The three phases can form a cyclical model, and reflections in each phase can have an effect on the next. In our study, when navigating self affects, participants conducted *Self-Observation* reflections at the *performance* phase and *Self-Judgement* reflections at the *self-reflection* phase (RQ1). Therefore, our findings show several promising designing opportunities to improve learners' self-regulated learning, and subsequently, learning outcomes.

6.1.1 Guiding Cyclical Reflections. Our findings show that among 212 collected reflection segments, 21% were ended by *Self-Observation* without proceeding to *Self-Judgment* (RQ1). According to prior literature, *self-observation* is insufficient, and people need to proceed to *self-judgment* to facilitate behavior changes [66, 93]. To improve self-regulated learning, multiple reflections should happen at different phases [66]. Future self-regulated tools could incorporate an intelligent agent to guide learners to conduct reflection processes in subsequent phases, e.g., *Self-Judgement*. In fact, conversational AI has been widely studied for providing guidance [23, 48]. Research shows that conversational AI can promote users' self-disclosure, which is also beneficial for users to reflect more on their own behaviors [48, 54]. Future systems may detect which affects learners are interacting with, apply NLP tools to recognize reflections in verbalized content [13], and then generate real-time responses to guide learners' reflections in the cyclical model.

Previous research found that it is challenging for students to be aware of and regulate emotions [72]. Our findings show that visualizing affective states could make learners' emotions more salient and support learners to reflect on. Self-Regulated learning tools should consider key processes of affects, such as the changes of their affects [4]. Some participants shared that our current design may not be efficient for learners to manually identify the dynamics of their affects as a result of information overloading (RQ2). Therefore, future self-regulated learning tools could explore more designs that can visualize changes or trends of learners' affective states; then automatically mark-up valuable points for learners to reflect on. For example, automated inferences on trends of personal data (e.g., exercising) may be incorporated in the design to improve learners' reflection [18].

6.1.2 Mitigating Confirmation Biases. Our findings also show that more interactions on negative learning timestamps were associated with more types of interactions with the tool, while such association was not found on positive learning timestamps (RQ1). It seemed that participants tended to agree with or prioritize results that made them feel good, and some even ignored negative and low-engagement results (RQ3). For example, P2 subconsciously searched for positive results in the AI graph and had blind-spots on the negative affects. Similar 'biases' and 'mistrust' had been observed in AI-assisted decision making scenarios, e.g. recent works showed that people often relied on AI-assisted decision support tools and tended to accept an AI suggestions even when the suggestions were wrong [16] [86]. This result suggests that future AI for self-regulated learning could detect what affects that the learners interact with, and try to mitigate potential confirmation bias, as well as guide learners to conduct more conscious observations. For example, some design let learners conduct reflection without AI first and have AI intervene at a later time [16].

Further on, individuals often have blind spots during reasoning [37], and viewing others' reasoning process could help identify overlooked information [91]. Thus, it is suggest to design tools that can support individuals to discern the mental states of others and to reflect on their own decision making process [22]. Collaborative reflection is considered as a feasible approach to foster more comprehensive reflections, as comparing individuals' data at different times and with peers could promote different sensemaking processes [55]. Therefore, future systems may look into how *Mirror* could support effective collaborative reflections. Comparing these affective states also require scalable solutions, which can visualize large numbers of learners' affective states both for peers and teachers.

6.1.3 Advancing Computer-supported Collaborative Learning. Previous research found that hearing-impaired and non-native speakers may not be able to understand video content when closed captions have errors, and they rely on other users (e.g., native speakers) to correct video captions by request [41]. *Mirror* can automatically detect learners confusion to augment the accessbility of video-based learning. For example, "confusion" of hearing-impaired and non-native language learners can be recognized by *Mirror*, such that timely requests can be sent to their peers or instructors for help. Also, previous works showed that teachers and learners, without accessibility needs, found non-verbal cues to be less noticeable for online teaching than in-person [17]. For video-based communications (e.g., [34, 44, 75]), *Mirror* can enrich learning statistics by adding affects.

6.2 Ethical Designs of AI-based Tools for Affective-Aware Learning

Below, we reflect on ethical considerations for designing learning analytic tools with affective awareness *Mirror*. Although *Mirror* recognizes facial movements of learners, further tools applying other AI recognition of affects may also benefit from our findings. For example, apply affective-aware tools in design sessions to assist analysis of usability testing [32].

6.2.1 Augmenting Emotion Regulation. Our findings revealed that learners had multiple emotion regulation when reviewing their own affects [2] – *acceptance* (e.g. accepting own affects fully), *problem solving* (e.g. rewatching a video clip), *reappraisal* (e.g. referring to self-report and video selfie). Meanwhile, a misalignment between continuous AI results and discrete self-reports interrupts their reflection processes and could result in a continuous loop of learners' ruminations with their affects (RQ2). The process aligns with destructive emotion regulation routines [2] – *rumination.* AI that augments the users' existing routines, and yet remains unobtrusive is referred as *unremarkable AI.* Users are more likely to encounter and prefer *unremarkable AI* since it naturally introduces predictions into their existing practices than incorporating AI that drastically changes their decision-making process [90]. Our research indicates that *Mirror* could also be *unremarkable* as it supports existing emotion regulation routines naturally. If the tool did not well support existing routines such as informing further actions, learners might have reactivity behaviors and find AI annoying (RQ3). Because monitoring, evaluation, and measuring technologies can produce reactivity behaviors if designed without a good understanding of existing social routines and systems [78].

6.2.2 Procedurality of Reflecting with AI-based Tools. Some participants pointed out that they could not only take the AI results as they were, and they would like to explore different data sources to reflect on the AI results (RQ3). Specifically, participants used the tool to map their recorded facial expressions with the AI results and to contextualize their reflections. In prior work, one researcher considered the ethics of AI-enabled learning tools regarding accuracy and learning assessment: *"classifying students in terms of educational tests has to consider the inherent ambiguity and variability in the measures, but for computer scientists, we usually consider them as rigid labels."* [38]. It was highlighted that machine learning models often fall into a formalism trap and fail to account for the full meanings of social concepts, which can be procedural, contextual, and contestable. To avoid the formalism trap, previous work recommended that real users should have the power to shape the technology, e.g. with interpretive flexibility [78]. In our findings, reflection with AI is procedural , and different data sources empower users to interpret the AI results (RQ3). Therefore, affective-aware tools should support interpretive flexibility. For example, systems can sense whether users' trust increases over time and give users more flexibility to interpret AI results when trust decreases.

6.2.3 Communicating Imperfect AI Results. *Mirror* presented the AI results in a fine granularity, which gave participants more agency in handling AI imperfectness (RQ3); however, it was also associated with information overloading (RQ2). More granular presentation of data is found to generate information overloading when using learning analytics tools [64] and health tracking tools [46]. We suggest that future affect-aware AI tools should better communicate imperfectness of the results. Prior works on explainable AI found result granularity should be small enough to inform users of its uncertainty without overwhelming them with too many details [8], and if the AI behaved unexpectedly or erroneously, users should be given the power of debugging to identify the offending fault and take control to make corrections [86]. Similar to other explainable AI domains, e.g. body scan in health domain [90], AI in education has larger volumes of data collected and is highly personalized [38]. Effective designs should communicate uncertainty in both directions: informing users of recognized results and correcting the results. For instance, system may provide several data in time-series, including change points detected [82], and allow learners to select the most preferred ones. Further research may explore how to better support learners' reasoning, e.g. by highlighting ambiguous cases along with text explanations of the AI results [76].

6.2.4 Transparency and Intelligibility of AI-Informed Decision-Making. Our findings (RQ3) provide empirical insights for enabling transparency and intelligibility of AI-informed decision-making, two important ethical issues of applying AI in the field of education [38]. For example, when AI applied for education, data ownership becomes problematic [38]. Beyond self-regulated learning online, such tools deployed in offline learning environments may create additional pressure and stress for students [58], and teachers may use data-driven tools for surveillance in online settings [57].

Privacy concerns of using AI-based tools for learning have been extensively discussed [12]. In our study, the participants did not express any privacy concerns of using *Mirror* though, probably because the automatically detected affects were only visible to themselves without sharing with others. However, future research should examine whether and how different ethical issues may impede using affective-aware tools for learning. Although we did not have participants with accessibility needs, previous research has found users with accessibility needs might have different facial expression features from general learners. For example, previous research found that users with autism spectrum disorders express their emotions with facial movement differently from general users [84]. Deaf and hard of hearing users' facial movement are better recognition via commercial recognition APIs compared to general users [81]. More research is needed to examine how facial recognition algorithms could be ethnically applied inclusively among diverse populations.

6.3 Limitations anf Future Work

The current tool is only designed for users after they finish watching videos and for them to review their affects. The tool recorded participants' video selfies but did not provide real-time detection of their affects. Also, the video selfies were only for participants' self reflection without sharing with others; therefore, the participants' ethical concerns on sharing the affects with other stakeholders (e.g., instructors or peers) may not reflect their real concerns, e.g., direct comparison between learner-learner and learner-crowd (e.g., [27]). The study only evaluated one video that generated limited types of affects. We plan to evaluate with more videos that are used for formal and informal education. Also, our findings are based on the current facial expression recognition technologies. When the accuracy of the algorithm improves, findings related to Human-AI interaction might be impacted and require further validation [69]. Our current study is about self-regulated learning, which doesn't always have quizzes. Further works will examine if using *Mirror* for reflections would impact learning outcomes, e.g., recalls of video content. Additionally, affective states can be used to trigger questions to facilitate learners' understanding of different concepts.

7 CONCLUSIONS

In this work, we designed, implemented, and evaluated a novel self-regulated learning tool, *Mirror*, that is enabled by facial recognition AI to support learning's reflection in video-based learning. We first conducted a need-finding study with 12 participants, which allowed us to gain insights on designing visualizations and interactions for reflections on Valence-Arousal of facial recognition results. Given the need-finding results, we then implemented *Mirror* and evaluated the tool with 16 participants. Our findings showed that *Mirror* supported participants' multiple reflection processes, and participants' affective valence (negative and positive) had an impact on the type of reflections participants conducted. Additionally, several technological and psychological factors impacted participants' perceived user experience of *Mirror*. For example, participants were be more willing to reflect on AI recognized affects that inform direct actions (e.g., re-watching video clips). The novel design of *Mirror* and the new findings make important implications for designing AI-based tools for self-regulated learning.

ACKNOWLEDGMENTS

This material is based upon work supported by the National Science Foundation under Grant No. 2119589. 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. This project was also made possible in part by the Institute of Museum and Library Services LG-70-18-0088-18. The views, findings, conclusions or recommendations expressed in this article do not necessarily represent those of the Institute of Museum and Library Services.

REFERENCES

[1] Hind Al-Fadda and Maha Al-Yahya. 2010. Using web blogs as a tool to encourage pre-class reading, post-class reflections and collaboration in higher education. *Online Submission* 7, 7 (2010), 100–106.

[2] Amelia Aldao, Susan Nolen-Hoeksema, and Susanne Schweizer. 2010. Emotion-regulation strategies across psychopathology: A meta-analytic review. *Clinical psychology review* 30, 2 (2010), 217–237.

[3] Roger Azevedo, Jason Harley, Gregory Trevors, Melissa Duffy, Reza Feyzi-Behnagh, François Bouchet, and Ronald Landis. 2013. Using trace data to examine the complex roles of cognitive, metacognitive, and emotional self-regulatory processes during learning with multi-agent systems. In *International handbook of metacognition and learning technologies.* Springer, 427–449.

[4] 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.

[5] Ebrahim Babaei, Namrata Srivastava, Joshua Newn, Qiushi Zhou, Tilman Dingler, and Eduardo Velloso. 2020. *Faces of Focus: A Study on the Facial Cues of Attentional States.* Association for Computing Machinery, New York, NY, USA, 1–13. https://doi.org/10.1145/3313831.3376566

[6] 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.

[7] 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.

[8] Umang Bhatt, Javier Antorán, Yunfeng Zhang, Q. Vera Liao, Prasanna Sattigeri, Riccardo Fogliato, Gabrielle Melançon, Ranganath Krishnan, Jason Stanley, Omesh Tickoo, Lama Nachman, Rumi Chunara, Madhulika Srikumar, Adrian Weller, and Alice Xiang. 2021. *Uncertainty as a Form of Transparency: Measuring, Communicating, and Using Uncertainty.* Association for Computing Machinery, New York, NY, USA, 401–413. https://doi.org/10.1145/3461702.3462571

[9] Robert Bodily and Katrien Verbert. 2017. Review of research on student-facing learning analytics dashboards and educational recommender systems. *IEEE Transactions on Learning Technologies* 10, 4 (2017), 405–418.

[10] Kirsten Boehner, Rogério DePaula, Paul Dourish, and Phoebe Sengers. 2007. How emotion is made and measured. *International Journal of Human-Computer Studies* 65, 4 (2007), 275–291.

[11] Monique Boekaerts. 1999. Self-regulated learning: Where we are today. *International journal of educational research* 31, 6 (1999), 445–457.

[12] Jason Borenstein and Ayanna Howard. 2021. Emerging challenges in AI and the need for AI ethics education. *AI and Ethics* 1, 1 (2021), 61–65.

[13] Nigel Bosch, Yingbin Zhang, Luc Paquette, Ryan Baker, Jaclyn Ocumpaugh, and Gautam Biswas. 2021. Students' Verbalized Metacognition During Computerized Learning *(CHI '21).* Association for Computing Machinery, New York, NY, USA, Article 680, 12 pages. https://doi.org/10.1145/3411764.3445809

[14] Kevin W Bowyer. 2004. Face recognition technology: security versus privacy. *IEEE Technology and society magazine* 23, 1 (2004), 9–19.

[15] 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.

[16] 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. *Proc. ACM Hum.-Comput. Interact.* 5, CSCW1, Article 188 (apr 2021), 21 pages. https://doi.org/10.1145/3449287

[17] Xinyue Chen, Si Chen, Xu Wang, and Yun Huang. 2021. "I Was Afraid, but Now I Enjoy Being a Streamer!": Understanding the Challenges and Prospects of Using Live Streaming for Online Education. *Proc. ACM Hum.-Comput. Interact.* 4, CSCW3, Article 237 (jan 2021), 32 pages. https://doi.org/10.1145/3432936

[18] 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* (Barcelona, Spain) *(PervasiveHealth '17)*. Association for Computing Machinery, New York, NY, USA, 173–182. https://doi.org/10.1145/3154862.3154881

[19] Bo Ae Chun and Hae Ja Heo. 2018. The Effect of Flipped Learning on Academic Performance as an Innovative Method for Overcoming Ebbinghaus' Forgetting Curve. In *Proceedings of the 6th International Conference on Information and Education Technology* (Osaka, Japan) *(ICIET '18)*. Association for Computing Machinery, New York, NY, USA, 56–60. https://doi.org/10.1145/3178158.3178206

[20] Ian Clark. 2012. Formative assessment: Assessment is for self-regulated learning. *Educational Psychology Review* 24, 2 (2012), 205–249.

[21] Ruth Cobos, Silvia Gil, Angel Lareo, and Francisco A. Vargas. 2016. Open-DLAs: An Open Dashboard for Learning Analytics. In *Proceedings of the Third (2016) ACM Conference on Learning @ Scale* (Edinburgh, Scotland, UK) *(L@S '16)*. Association for Computing Machinery, New York, NY, USA, 265–268. https://doi.org/10.1145/2876034.2893430

[22] Kym Craig, Daniel Hale, Catherine Grainger, and Mary E Stewart. 2020. Evaluating metacognitive self-reports: systematic reviews of the value of self-report in metacognitive research. *Metacognition and Learning* 15, 2 (2020), 155–213.

[23] Marilyn Wood Daudelin. 1996. Learning from experience through reflection. *Organizational dynamics* 24, 3 (1996), 36–48.

[24] 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.

[25] Didan Deng, Zhaokang Chen, Yuqian Zhou, and Bertram Shi. 2020. Mimamo net: Integrating micro-and macro-motion for video emotion recognition. In *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 34. 2621–2628.

[26] Anna Dipace, Bojan Fazlagic, and Tommaso Minerva. 2019. The design of a learning analytics dashboard: EduOpen MOOC platform redefinition procedures. *Journal of E-learning and Knowledge Society* 15, 3 (2019), 29–47.

[27] Hendrik Drachsler and Wolfgang Greller. 2016. Privacy and Analytics: It's a DELICATE Issue a Checklist for Trusted Learning Analytics. In *Proceedings of the Sixth International Conference on Learning Analytics & Knowledge* (Edinburgh, United Kingdom) *(LAK '16)*. Association for Computing Machinery, New York, NY, USA, 89–98. https://doi.org/10.1145/2883851.2883893

[28] Eva Durall and Begoña Gros. 2014. Learning Analytics as a Metacognitive Tool.. In *CSEDU (1)*. 380–384.

[29] Paul Ekman and Wallace V Friesen. 1976. Measuring facial movement. *Environmental psychology and nonverbal behavior* 1, 1 (1976), 56–75.

[30] Paul Ed Ekman and Richard J Davidson. 1994. *The nature of emotion: Fundamental questions.* Oxford University Press.

[31] 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.

[32] Mingming Fan, Xianyou Yang, TszTung Yu, Q. Vera Liao, and Jian Zhao. 2022. Human-AI Collaboration for UX Evaluation: Effects of Explanation and Synchronization. *Proc. ACM Hum.-Comput. Interact.* 6, CSCW1, Article 96 (apr 2022), 32 pages. https://doi.org/10.1145/3512943

[33] David García-Solórzano, Eugènia Santamaría, Jose A Moran, Germán Cobo, and Carlos Monzo. 2018. Personal informatics systems for supporting self-regulation in online learning environments. *Computer Applications in Engineering Education* 26, 5 (2018), 1445–1456.

[34] Jens Emil Grønbæk, Banu Saatçi, Carla F. Griggio, and Clemens Nylandsted Klokmose. 2021. MirrorBlender: Supporting Hybrid Meetings with a Malleable Video-Conferencing System. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems* (Yokohama, Japan) *(CHI '21)*. Association for Computing Machinery, New York, NY, USA, Article 451, 13 pages. https://doi.org/10.1145/3411764.3445698

[35] Philip J. Guo, Juho Kim, and Rob Rubin. 2014. How Video Production Affects Student Engagement: An Empirical Study of MOOC Videos. In *Proceedings of the First ACM Conference on Learning @ Scale Conference* (Atlanta, Georgia, USA) *(L@S '14)*. Association for Computing Machinery, New York, NY, USA, 41–50. https://doi.org/10.1145/2556325.2566239

[36] Jason M Harley, François Bouchet, and Roger Azevedo. 2013. Aligning and comparing data on emotions experienced during learning with MetaTutor. In *International Conference on Artificial Intelligence in Education*. Springer, 61–70.

[37] Richards J Heuer. 1999. *Psychology of intelligence analysis.* Center for the Study of Intelligence.

[38] 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.

[39] 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.

[40] Silas Hsu, Tiffany Wenting Li, Zhilin Zhang, Max Fowler, Craig Zilles, and Karrie Karahalios. 2021. Attitudes Surrounding an Imperfect AI Autograder. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems* (Yokohama, Japan) *(CHI '21)*. Association for Computing Machinery, New York, NY, USA, Article 681, 15 pages. https://doi.org/10.1145/3411764.3445424

[41] Yun Huang, Yifeng Huang, Na Xue, and Jeffrey P. Bigham. 2017. Leveraging Complementary Contributions of Different Workers for Efficient Crowdsourcing of Video Captions *(CHI '17)*. Association for Computing Machinery, New York, NY, USA, 4617–4626. https://doi.org/10.1145/3025453.3026032

[42] 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.

[43] Ioana Jivet, Maren Scheffel, Marcel Schmitz, Stefan Robbers, Marcus Specht, and Hendrik Drachsler. 2020. From students with love: An empirical study on learner goals, self-regulated learning and sense-making of learning analytics in higher education. *The Internet and Higher Education* 47 (2020), 100758.

[44] Matthew K. Miller, Martin Johannes Dechant, and Regan L. Mandryk. 2021. Meeting You, Seeing Me: The Role of Social Anxiety, Visual Feedback, and Interface Layout in a Get-to-Know-You Task via Video Chat.. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems* (Yokohama, Japan) *(CHI '21)*. Association for Computing Machinery, New York, NY, USA, Article 339, 14 pages. https://doi.org/10.1145/3411764.3445664

[45] Dongho Kim, Il-Hyun Jo, Donggil Song, Hua Zheng, Jingwei Li, Jiawen Zhu, Xing Huang, Wei Yan, and Zhen Xu. 2021. Self-regulated learning strategies and student video engagement trajectory in a video-based asynchronous online course: a Bayesian latent growth modeling approach. *Asia Pacific Education Review* (2021), 1–13.

[46] Yoojung Kim, Bongshin Lee, and Eun Kyoung Choe. 2019. Investigating data accessibility of personal health apps. *Journal of the American Medical Informatics Association* 26, 5 (2019), 412–419.

[47] René F. Kizilcec. 2016. How Much Information? Effects of Transparency on Trust in an Algorithmic Interface. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems* (San Jose, California, USA) *(CHI '16)*. Association for Computing Machinery, New York, NY, USA, 2390–2395. https://doi.org/10.1145/2858036.2858402

[48] 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

[49] Dimitrios Kollias, Attila Schulc, Elnar Hajiyev, and Stefanos Zafeiriou. 2020. Analysing affective behavior in the first abaw 2020 competition. In *2020 15th IEEE International Conference on Automatic Face and Gesture Recognition (FG 2020)*. IEEE, 637–643.

[50] Barry Kort, Rob Reilly, and Rosalind W Picard. 2001. An affective model of interplay between emotions and learning: Reengineering educational pedagogy-building a learning companion. In *Proceedings IEEE International Conference on Advanced Learning Technologies*. IEEE, 43–46.

[51] 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.

[52] 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.

[53] Jyoti Kumari, R Rajesh, and KM Pooja. 2015. Facial expression recognition: A survey. *Procedia computer science* 58 (2015), 486–491.

[54] Yi-Chieh Lee, Naomi Yamashita, Yun Huang, and Wai Fu. 2020. "I Hear You, I Feel You": Encouraging Deep Self-Disclosure through a Chatbot. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. Association for Computing Machinery, New York, NY, USA, 1–12. https://doi.org/10.1145/3313831.3376175

[55] Lisa Lim, Shane Dawson, Srecko Joksimovic, and Dragan Gašević. 2019. Exploring Students' Sensemaking of Learning Analytics Dashboards: Does Frame of Reference Make a Difference?. In *Proceedings of the 9th International Conference on Learning Analytics & Knowledge* (Tempe, AZ, USA) *(LAK19)*. Association for Computing Machinery, New York, NY, USA, 250–259. https://doi.org/10.1145/3303772.3303804

[56] Lisa-Angelique Lim, Dragan Gasevic, Wannisa Matcha, Nora'Ayu Ahmad Uzir, and Shane Dawson. 2021. Impact of Learning Analytics Feedback on Self-Regulated Learning: Triangulating Behavioural Logs with Students' Recall. In *LAK21: 11th International Learning Analytics and Knowledge Conference* (Irvine, CA, USA) *(LAK21)*. Association for Computing Machinery, New York, NY, USA, 364–374. https://doi.org/10.1145/3448139.3448174

[57] Alex Jiahong Lu, Tawanna R. Dillahunt, Gabriela Marcu, and Mark S. Ackerman. 2021. Data Work in Education: Enacting and Negotiating Care and Control in

Teachers' Use of Data-Driven Classroom Surveillance Technology. *Proc. ACM Hum.-Comput. Interact.* 5, CSCW2, Article 452 (oct 2021), 26 pages. https://doi.org/10.1145/3479596

[58] Vidushi Marda and Shazeda Ahmed. 2021. Emotional Entanglement: China's emotion recognition market and its implications for human rights.

[59] Michael E Martinez. 2006. What is metacognition? *Phi delta kappan* 87, 9 (2006), 696–699.

[60] Wannisa Matcha, Dragan Gašević, Abelardo Pardo, et al. 2019. A systematic review of empirical studies on learning analytics dashboards: A self-regulated learning perspective. *IEEE Transactions on Learning Technologies* 13, 2 (2019), 226–245.

[61] Barbara L McCombs. 1989. Self-regulated learning and academic achievement: A phenomenological view. In *Self-regulated learning and academic achievement*. Springer, 51–82.

[62] Atasi Mohanty et al. 2016. Affective pedagogical agent in e-learning environment: A reflective analysis. *Creative Education* 7, 04 (2016), 586.

[63] Ali Mollahosseini, Behzad Hasani, and Mohammad H Mahoor. 2017. Affectnet: A database for facial expression, valence, and arousal computing in the wild. *IEEE Transactions on Affective Computing* 10, 1 (2017), 18–31.

[64] Andy Nguyen, Tuure Tuunanen, Lesley Gardner, and Don Sheridan. 2021. Design principles for learning analytics information systems in higher education. *European Journal of Information Systems* 30, 5 (2021), 541–568.

[65] 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.

[66] Ernesto Panadero. 2017. A review of self-regulated learning: Six models and four directions for research. *Frontiers in psychology* 8 (2017), 422.

[67] Zacharoula Papamitsiou and Anastasios A Economides. 2015. Temporal learning analytics visualizations for increasing awareness during assessment. *International Journal of Educational Technology in Higher Education* 12, 3 (2015), 129–147.

[68] Yeonjeong Park and I-H Jo. 2015. Development of the learning analytics dashboard to support students' learning performance. *Journal of Universal Computer Science* 21, 1 (2015), 110.

[69] Arun Rai. 2020. Explainable AI: From black box to glass box. *Journal of the Academy of Marketing Science* 48, 1 (2020), 137–141.

[70] Michael J Reiss. 2021. The use of AI in education: Practicalities and ethical considerations. *London Review of Education* (2021).

[71] Nancy A Remington, Leandre R Fabrigar, and Penny S Visser. 2000. Reexamining the circumplex model of affect. *Journal of personality and social psychology* 79, 2 (2000), 286.

[72] Anna D Rowe and Julie Fitness. 2018. Understanding the role of negative emotions in adult learning and achievement: A social functional perspective. *Behavioral sciences* 8, 2 (2018), 27.

[73] James A Russell. 1980. A circumplex model of affect. *Journal of personality and social psychology* 39, 6 (1980), 1161.

[74] Anas Samara, Leo Galway, Raymond Bond, and Hui Wang. 2019. Affective state detection via facial expression analysis within a human–computer interaction context. *Journal of Ambient Intelligence and Humanized Computing* 10, 6 (2019), 2175–2184.

[75] 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 *(CHI '21)*. Association for Computing Machinery, New York, NY, USA, Article 252, 13 pages. https://doi.org/10.1145/3411764.3445615

[76] Mike Schaekermann, Graeme Beaton, Elaheh Sanoubari, Andrew Lim, Kate Larson, and Edith Law. 2020. *Ambiguity-Aware AI Assistants for Medical Data Analysis*. Association for Computing Machinery, New York, NY, USA, 1–14. https://doi.org/10.1145/3313831.3376506

[77] Dale H Schunk. 2008. Metacognition, self-regulation, and self-regulated learning: Research recommendations. *Educational psychology review* 20, 4 (2008), 463–467.

[78] Andrew D. Selbst, Danah Boyd, Sorelle A. Friedler, Suresh Venkatasubramanian, and Janet Vertesi. 2019. Fairness and Abstraction in Sociotechnical Systems. In *Proceedings of the Conference on Fairness, Accountability, and Transparency* (Atlanta, GA, USA) *(FAT* '19)*. Association for Computing Machinery, New York, NY, USA, 59–68. https://doi.org/10.1145/3287560.3287598

[79] Neil Selwyn. 2020. Re-imagining 'Learning Analytics'… a case for starting again? *The Internet and Higher Education* 46 (2020), 100745.

[80] Nathan Semertzidis, Michaela Scary, Josh Andres, Brahmi Dwivedi, Yutika Chandrashekhar Kulwe, Fabio Zambetta, and Florian Floyd Mueller. 2020. *Neo-Noumena: Augmenting Emotion Communication*. Association for Computing Machinery, New York, NY, USA, 1–13. https://doi.org/10.1145/3313831.3376599

[81] Irene Rogan Shaffer. 2018. Exploring the Performance of Facial Expression Recognition Technologies on Deaf Adults and Their Children. In *Proceedings of the 20th International ACM SIGACCESS Conference on Computers and Accessibility* (Galway, Ireland) *(ASSETS '18)*. Association for Computing Machinery, New York, NY, USA, 474–476. https://doi.org/10.1145/3234695.3240986

[82] Charles Truong, Laurent Oudre, and Nicolas Vayatis. 2020. Selective review of offline change point detection methods. *Signal Processing* 167 (2020), 107299.

[83] Katrien Verbert, Erik Duval, Joris Klerkx, Sten Govaerts, and José Luis Santos. 2013. Learning analytics dashboard applications. *American Behavioral Scientist* 57, 10 (2013), 1500–1509.

[84] Simon Wallace, Michael Coleman, and Anthony Bailey. 2008. An investigation of basic facial expression recognition in autism spectrum disorders. *Cognition and Emotion* 22, 7 (2008), 1353–1380.

[85] Bryan Wang, Meng Yu Yang, and Tovi Grossman. 2021. Soloist: Generating Mixed-Initiative Tutorials from Existing Guitar Instructional Videos Through Audio Processing. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems* (Yokohama, Japan) *(CHI '21)*. Association for Computing Machinery, New York, NY, USA, Article 98, 14 pages. https://doi.org/10.1145/3411764.3445162

[86] Danding Wang, Qian Yang, Ashraf Abdul, and Brian Y. Lim. 2019. Designing Theory-Driven User-Centric Explainable AI. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (Glasgow, Scotland Uk) *(CHI '19)*. Association for Computing Machinery, New York, NY, USA, 1–15. https://doi.org/10.1145/3290605.3300831

[87] 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.

[88] Kimberly Williamson and Rene Kizilcec. 2022. A Review of Learning Analytics Dashboard Research in Higher Education: Implications for Justice, Equity, Diversity, and Inclusion. In *LAK22: 12th International Learning Analytics and Knowledge Conference* (Online, USA) *(LAK22)*. Association for Computing Machinery, New York, NY, USA, 260–270. https://doi.org/10.1145/3506860.3506900

[89] Alyssa Friend Wise, Juan Pablo Sarmiento, and Maurice Boothe Jr. 2021. Subversive Learning Analytics. In *LAK21: 11th International Learning Analytics and Knowledge Conference* (Irvine, CA, USA) *(LAK21)*. Association for Computing Machinery, New York, NY, USA, 639–645. https://doi.org/10.1145/3448139.3448210

[90] Qian Yang, Aaron Steinfeld, and John Zimmerman. 2019. Unremarkable AI: Fitting Intelligent Decision Support into Critical, Clinical Decision-Making Processes. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (Glasgow, Scotland Uk) *(CHI '19)*. Association for Computing Machinery, New York, NY, USA, 1–11. https://doi.org/10.1145/3290605.3300468

[91] Chi-Hsien (Eric) Yen, Haocong Cheng, Grace Yu-Chun Yen, Brian P. Bailey, and Yun Huang. 2021. Narratives + Diagrams: An Integrated Approach for Externalizing and Sharing People's Causal Beliefs. *Proc. ACM Hum.-Comput. Interact.* 5, CSCW2, Article 444 (oct 2021), 27 pages. https://doi.org/10.1145/3479588

[92] Kaipeng Zhang, Zhanpeng Zhang, Zhifeng Li, and Yu Qiao. 2016. Joint face detection and alignment using multitask cascaded convolutional networks. *IEEE Signal Processing Letters* 23, 10 (2016), 1499–1503.

[93] Barry J Zimmerman and Adam R Moylan. 2009. Self-regulation: Where metacognition and motivation intersect. In *Handbook of metacognition in education*. Routledge, 311–328.