Revealing the Hidden Structure of Affective States During Emotion Regulation in Synchronous Online Collaborative Learning

Belle Dang
University of Oulu, Finland
Huong.dang@student.oulu.fi

Andy Nguyen
University of Oulu, Finland
Andy.nguyen@oulu.fi

Yvonne Hong
Victoria University of Wellington, New Zealand
Yvonne.hong@vuw.ac.nz

Bich-Phuong Thi Nguyen
VNU Hanoi University of Languages and International Studies, Vietnam
bichphuong@vnu.edu.vn

Bao-Nhi Dang Tran
RMIT University, Vietnam
s3751881@rmit.edu.vn

Abstract

This study aims to explore the use of advanced technologies such as artificial intelligence (AI) to reveal learners' emotion regulation. In particular, this study attempts to discover the hidden structure of affective states associated with facial expression during challenges, interactions, and strategies for emotion regulation in the context of synchronous online collaborative learning. The participants consist of 18 higher education students (N=18) who worked collaboratively in groups. The Hidden Markov Model (HMM) results indicated interesting transition patterns of latent state of emotion and provided insights into how learners engage in the emotion regulation process. This study demonstrates a new opportunity for theoretical and methodology advancement in the exploration of AI in researching socially shared regulation in collaborative learning.

Keywords: Facial expression recognition, hidden Markov model, computer-supported collaborative learning, socially shared regulation, emotion regulation

1. Introduction

Self-regulated learning (SRL) has been recognized as an essential contributor to individual learning success, while co-regulation (CoRL) and socially shared regulation (SSRL) significantly impact collaborative learning success. SRL involves a cyclical process of monitoring, adapting, and reflecting on the behavioural, cognitive, metacognitive, motivational, and emotional aspects of learning towards achieving learning goals. In collaborative learning, CoRL and SSRL, through social interactions, reflect the nature of group work with the iterative deliberate, strategic, and transactive planning, task execution, and reflection of a group (Järvelä et al., 2018). Promoting learning regulation (SRL, CoRL, and SSRL) would not only leverage the learners' academic performance but also benefit their lifelong success (Xiao & Yang, 2019). However, it is challenging to detect and support learning regulation due to the unobservable nature of emotional and cognitive processes at its core (Järvelä & Bannert, 2021).

Fortunately, recent technological advancement has allowed for the collection and analysis of new data channels to better understand the learning processes (Dindar et al., 2022; Nguyen et al., 2022). Contemporary research has attempted to utilize advanced technologies such as physiological sensors and artificial intelligence (AI) to provide novel insights into learning regulation processes. For instance, Dindar et al. (2022) have proposed an approach to measure shared physiological arousal events in collaborative problem-solving. Another example is the AI deep learning model developed by Nguyen et al. (2022) that evicted the use of advanced technologies to detect regulatory activities in collaborative learning automatically. Following this line of research, this study attempts to utilize AI facial recognition and machine learning modelling to reveal the hidden structure of learners' facial emotion states during emotion regulation in synchronous online collaborative learning. Particularly, this paper addresses the following research questions:

1) What types of hidden affective states can be identified based on facial expression recognition?

URI: https://hdl.handle.net/10125/102632
978-0-9981331-6-4

2) How do the hidden affective states associate with regulatory challenges, interactions, and strategies for emotion regulation in synchronous online collaborative learning?

2. Theoretical Background

2.1. Socially Shared Regulation of Learning

Historically, learning regulation is portrayed as an individualized cognitive-constructive process whereby emphasis is placed on learners' deliberate adaptation in regulating their learning by strategizing their learning plans, goal setting, monitoring, and regulation of cognitive, motivational, emotional, and behavioural processes towards achieving a particular outcome. The field of study has since evolved significantly with contemporary learning theories emphasizing active constructions of knowledge and the notion of shared knowledge construction, whereby self-regulation does not always occur independently and can be originated or imposed on by others (Hadwin et al., 2015). This has since resulted in a range of learning regulatory models along the continuum ranging from individual constructivist (i.e. self-regulatory learning) to social-constructivist perspectives of learning (i.e. socially shared regulation of learning) models (Hadwin et al., 2005).

Socially Shared Regulation of Learning (SSRL) refers to the collective regulatory activities of multiple individuals in a group setting, either in-person or in an online learning environment. It is stated that simply putting learners into groups for collaborative activities does not automatically guarantee learning success. Instead, to succeed in the construction of new knowledge, learners need to negotiate and align shared goals and iteratively fine-tune shared metacognition, behaviour, emotion, and motivation conditions collectively towards a shared goal (Hadwin et al., 2015; Isohätälä et al., 2017; Järvelä et al., 2019).

The interest in understanding SSRL has increased over the last decade. Social interactions in collaborative settings invite learners to share and debate differing perspectives, extending their knowledge beyond their innate capabilities. While empirical findings have heavily corroborated the benefits of self-regulated learning, there is still a dearth of research in capturing, for instance, how these regulatory processes beyond the individual emerge in a collaborative learning context (Hadwin et al., 2015; Malmberg et al., 2017).

2.2. Methods and challenges in studying SSRL

Learning regulation processes are challenging to measure, which brings the need to transition from traditional methods to multi-methodological approaches to capture both objective and subjective traces of the regulatory processes. A systematic review of SSRL by Järvelä et al. (2019) stated that whilst there has been methodological progress in the field, ongoing methodological development needs to address and capture more relevant markers to help understand the dynamic process of regulation of for instance, cognition, motivation, physiological states, and emotion attributes (Järvelä et al., 2019).

The first challenge noted is that learning regulation involves a cyclical adaptation that is challenging to capture. Individuals are constantly utilizing their metacognition to adapt their learning strategically, and these cycles of learning adaptation may vary across each cycle (Zimmerman, 2013). Next, the need to holistically comprehend and capture each learner's various and intertwined elements (e.g., emotion, motivation, cognition) and their interactions and regulation with others within the social learning context to capture SSRL authentically (Järvelä et al., 2019).

The challenges have since offset an evolution of data collection methods in self-regulatory learning. Considering the dynamic nature of learning regulation, retrospective subjective measures such as self-reported data from surveys and interviews are deemed insufficient in capturing the exact moments when those aspects of regulatory actions (e.g., cognition, motivation, physiological and emotional states) occur and how these influence each other. This increased emphasis on trace data or real-time measurements. Multimodal data, such as log files of time-stamped descriptions of observable interactions between learners and content, eye-tracking, screen recordings of interactions between learners and machines, think-aloud protocols, and physiological sensors, when implemented over time, are perceived to be more useful in providing objective insights into the patterns and changes in the regulatory processes with specific timeframes (Azevedo & Gašević, 2019; Järvelä et al., 2019).

Furthermore, gathering in-situ data on challenges faced by learners in authentic learning tasks is essential as this provides the opportunity to explore the unique makeup of learners and their interactions when challenges emerge, as well as what challenges faced in different social, technological, and contextual factors and trace the regulation as it evolves within a given scenario. This allows for rich multilayer data considerations such as objective data (e.g., physiological responses and eye-tracking) to be triangulated with subjective data (e.g., learners' perceptions and intent) to better understand the traces of regulatory behaviours and processes as temporally unfolding events that are contextualized in-situ (Hadwin et al., 2015; Järvelä et al., 2019).

While multimodal approaches with emerging technologies would propel the self-regulatory learning (SRL) field towards a higher interference of the learning process (Harley et al., 2015), more work needs to be done to increase the reliability and validity of the methodology. Holistic research methods are required to pinpoint the significance of which modalities reveal specific events and measures of the SSRL processes (Järvelä et al., 2019). Nevertheless, the findings were fragmented by conventional statistical and data mining techniques used to detect, measure and infer the complex and messy aspects of the learning regulation process. Leveraging artificial intelligence technology would help widen the multimodal data channels, increase the understanding of the complex processes by tracing and detecting more regulatory markers to augment advanced learning technologies to provide a more holistic, real-time, intelligent and personalized scaffolding and feedback according to each learner's regulatory needs (Azevedo & Gašević, 2019; Järvelä et al., 2019).

2.3. Artificial Intelligence for studying Emotions in SSRL

As aforementioned, with emotion, motivation, and cognitive processes being at the core of learning regulation, emerging technologies would play a vital role in capturing the intricacy of these regulatory processes (Azevedo & Gašević, 2019; Järvelä et al., 2019). While emotions were traditionally described as an indication of an individual's internal states, emotions are also attuned to interpersonal responses (Rogat & Adams- Wiggins, 2015). Emotions profoundly influence an individual's cognitive processes; for example, negative emotions would decrease learners' attention and interest, leading to demotivating learning behaviours (Tyng et al., 2017).

Furthermore, in a collaborative context, the emotional expression of a group member is shaped by the atmosphere of the group. For instance, individuals who received a hostile reaction, such as an unhappy or disagreeing facial expression, would feel rejected and, in turn, contribute less collaboratively. The social contagion of emotions amongst interacting learners effectively functions as a regulator in the transference of a positive or negative learning experience (Heerdink et al., 2013).

While the overall learning experiences are easily measured after the collaborative event, it is often difficult to capture and measure the fluidity of the short-term affective states of individuals. The different states of emotions are mimicked and shared amongst all group members through cycles of interactions (Rogat & Adams-Wiggins, 2015). Emotional mimicry, for instance, has been validated by studies as a marker of initial affiliative bond and empathy amongst learners. Nevertheless, the transient states of emotion following the learning tasks' progress and continual relations with others influence coordination and group cohesion. Understanding and capturing these temporal cycles of emotions and emotional contagion in a fine-grained manner would better help identify pain points in SSRL and timely prompts and cues to help regroup members and maintain the quality of the learning experience (Dindar et al., 2020).

A promising technique for identifying temporal and cyclical emotions in the collaborative learning context would be implementing a time-stamped video-based facial expression recognition method. The method, consisting of time-stamped frames, is useful in providing the fine-grained level of details (Dindar et al., 2020), for instance, the exact moment of changes in facial expressions, matching the specific tasks learners were working on. Furthermore, the analysis of the audio captured would provide a richer context for understanding the regulatory dynamics and social atmosphere of the group at a given point in time. However, current techniques to collect multimodal data are still somewhat lacking (Järvelä et al., 2019). The manual coding of facial expressions, speech, and gestures is not only highly time-consuming but also poses data triangulation and validity challenges (Graham et al., 2020).

Given that, there are now emerging technologies powered by Artificial Intelligence in assisting in multimodal data collection, with deep learning analysis and, in turn, real-time improved personalized and predictive abilities based on machine learning algorithms. AI technologies have the capability not only to increase the accuracy of the frame-by-frame analysis of emotions but, when integrated with physiological sensors (Zhang et al., 2020) and triangulated with audio data, would increase the depth of analysis. This results in learned and customized responses based on real-time identification of regulation and prompts to help regulate individuals' learning needs more effectively. Whilst AI solutions have been vastly implemented in other sectors, such as the health sector, the implementation of AI solutions for SRL research is still ongoing progress. Hence, with the implementation of AI-based facial recognition technology and an advanced analyzing algorithm, this study aims to better capture and understand emotions in SSRL and uncover the latent stages characterized by these emotions.

3. Methods

3.1. Data Collection

The participants in this study were 18 university students (age = 28), enrolled in an academic English course at a Vietnamese university. Their participation in this study was voluntary and accompanied by written informed consent. In addition, students received monetary compensation for their participation in the study.

As for the learning task, a group-based writing activity was carried out on Zoom, an online collaborative learning environment. The students were divided into groups of three and were sent into breakout rooms to discuss and complete a writing assignment. They were given a topic to discuss and were given thirty minutes to complete their writing report. The entire duration of their discussion and writing were recorded.

To minimize unexpected interference and technical problems, students were instructed to sit in a quiet environment and perform quality checks (i.e., headphones, internet connection, camera) before the group activity. During the collaborative session, they were given the option of sharing their computer screens. However, all participants' cameras had to be turned on. This was to ensure that their facial expressions, postures, and hand gestures were captured for the video-processed data. To assist students with their learning, supporting prompts were provided at the beginning of the session. These included, for example, information about task structure and strategies to form an outline. Apart from these, further support or feedback is not provided during the entire collaborative session to encourage students to proactively engage in socially shared regulation. A 200-word paragraph was produced as a result of the task and was assessed based on the official marking rubrics for the Common European Framework of Reference for Languages (CEFR).

3.2. Video: Qualitative Analysis

Video-based data of students' collaborative writing sessions were analyzed through qualitative coding and AI facial expression recognition (FER) analysis. The Hidden Markov Model (HMM) was adopted to fit the best latent structure in accordance with the affective state sequence, while quantitative statistical analysis of aligned video coding and HMM outputs informed the learners' emotions related to different regulatory activities.

A coding scheme was adapted from prior studies to examine how participants demonstrated their initiative in taking the lead and following behaviours as responses to verbal and non-verbal exchanges in the collaborative session (Järvenoja et al., 2019; Malmberg et al., 2017). Nonetheless, instead of analyzing the video recordings in 30-second segments, this research aims to apply a more detailed and sophisticated method. The 30-second segment video analysis has been criticized as being insufficient for the machine learning approach (Nguyen et al., 2022). In this coding method, a code is assigned to a group member's talking turn during the collaborative session. The primary focus of each group's interactions is coded; cognitive interactions, task execution, socio-emotional interactions shown in verbal, bodily, and emotionally charged indicators, and other non-task-related activities. Furthermore, as for the challenge types, the code is applied when participants demonstrated their cognitive difficulties in completing the task, emotional and motivational issues in regulating negative emotions, social context and interaction challenges in the surrounding environment, communication, and team collaboration. Last but not least, emotional regulation strategies were selected for coding when students show their mental encouragement and social reinforcement by creating a positive environment, task structuring to refocus task behaviour of distracted members, and raising awareness to support other team members in negative emotion management.

Participants' behaviour and how they showed interactions with other teammates during the collaborative session were coded using the combination of both video recordings and transcripts. After being briefed on the coding scheme, two independent researchers joined in the coding phase and piloted a video recording analysis to assess inter-rater consistency. Cohen's kappa coefficients of 0.71 (Interaction), 0.84 (Challenge), and 0.83 (Emotion Regulation Strategies) indicate that the emerged codes are highly reliable. The remaining video recordings were then coded individually by two researchers, while researcher-researcher corroboration was also maintained to discuss issues that arise throughout the coding procedure. The coding scheme for qualitative video analysis is shown in Table 1.

3.4. AI for Mining Facial Emotion

Emotion has been used in the past to investigate the mechanical properties of its role in learning and

Table 1. Coding scheme for video qualitative analysis

Categories	Description	Examples
Interaction		
Cognitive interaction	Interaction focuses on the learning-related higher mental process toward the metacognitive level (monitoring and controlling).	*S1: We need to research first!* *S2: Why should we for a B1 B2 English level?* *S1: In the document, it said the writing will be assessed based on C1 level, right?*
Socio-emotional interaction	Interaction focuses on strong expressions of socio-emotion with clear negative/positive affect nature (e.g., showing gratitude, approving, joking, disputing, criticizing). Expression included verbal and/or non-verbal indicators.	*S1: It kept flickering, so annoying!* *S2: It's too loud in my place, right? [expressing annoyance]* *S1: No, I couldn't hear anything*
Task execution interaction	Interaction that primary focus on carrying out task requirement, and completing the task: - Writing out the task - Read task instruction	*[Tying and read aloud] … CO2 and SO2…* *[Saying out loud] I am sending the docs link now.*
Other interaction	Interaction that is unrelated to the task topic or objective without a clear indicator of significant emotion. E.g., Talking about: - Out-of-school or school-related activities.	*S1: I can hear what Hang said echo all till here in Zoom.* *S2: Of course, being in the same room will have some differences.*
Regulation Challenges		
Cognitive challenge	Coded when group member(s) indicated difficulties related to higher learning mental processes such as memory processing, understanding task, finding solutions, ability to solve the task by choosing answers, strategies, etc.	*But those things don't have clear effect. It wouldn't make a strong point. [Stuck and doesn't know what to use instead]*
Emotional and motivational challenge	Coded when group member shows clear indicators of negative emotion, their inability to control them hinders the task progression. This included annoyance, boredom, lack of interest, motivation, etc.	*I am scared I am going to make a grammar mistake. [Hesitate to continue the task]* *We should write now or else we won't make it in time.*
Social context and other interaction challenges	Coded when group member experiences other types of difficulties that hinder the task progression. This includes environmental context such as resources-related issues or social context such as conflict of working style.	*Shall I turn off the shared screen, it's blocking my vision.*
Emotion Regulation Strategies		
Encouragement	Coded when group members provide emotional support to others or the groups by praising or supporting each other.	*Don't worry, this is like a 7.0 equal grade!* *Wow, that is good. Keep going.*
Social Reinforcement	Coded when group member tries to regulate others' negative emotion and the social atmosphere by highlighting and reassuring the positive aspect of their situation.	*[Members is wondering about the requirement of the task]* *S1: 230 words is a little bit long, but it is acceptable. It wouldn't cost the grade.*
Task Structuring	Coded when group member tries to draw the focus to task-related behaviours. This could be to avoid off-task behaviour that causes frustration to other members or to diffuse a potentially tense situation.	*[S1 and S2 were negotiating on which term to use and building up hostility]* *S3: Okay, so how about instead, we just talk in general and then add other criteria here?*
Increasing Awareness	Coded when group members attempt to get others to become more aware of their negative emotions or affect state, thereby facilitating the regulation of these feelings.	*S1: I think you are overthinking, and you don't have to worry to that extent…. if I were you, I think I would feel like I am trying to think about the effect of that sources…*

increase in self-regulation, yet there is still a lack of systematic understanding of how it unfolds and interacts within these invisible cognitive processes (Järvenoja et al., 2019).

To accurately identify the emotion exhibited by faces, it is important that we choose a model that not only possesses robust face detection for collaboration but is also effective when applied to the narrow scope of data collected. There have been many different methods for analyzing facial expressions in recent years, particularly since the resurgence of interest in convolutional neural networks. Automatic Facial Expression Recognition (FER) has been investigated extensively in the field of learning and teaching to recognize and track students' behaviours. The deep learning algorithms can not only detect human emotions directly from video recordings but also classify them on a continuous scale of valence (Khaireddin & Chen, 2021), which makes it possible to determine students' emotional states in sequences of recursive cognitive processes of self-regulation. The relatively small scope of the dataset in our synchronous online collaborative group project (less than 100 instances per class) makes it unlikely to train a FER model from scratch or even fine-tune an existing model. Therefore, imp2pose has been selected for use in the facial localization module as it has previously been trained on the Wider Face dataset and has shown a reliable performance of 3.9 mean square error (Zhu et al., 2016).

3.5. Hidden Markov Model

The Hidden Markov Model (HMM) is constructed by enhancing the Markov chain, a model that provides information on the probabilities of sequences of random variables or states, where each can take on values from a given set. Markov's assumption stated that the current state of the process is sufficient to forecast the future state of the process, and the prediction should be as accurate as when based on the process's past (Biswal et al., 2021). HMMs are models in which the distribution that generates an observation depends on the state of an underlying unobserved Markov process. The behaviour of a hidden Markov model is described along two dimensions: one indicated by "observable" and the other by "unobservable" or "hidden" (Coelho, 2019). In both models of existence, the state is the fundamental unit of processing. The states connected with the unobserved plane are referred to as hidden states, and those in the viewable portion are referred to as observable states. In HMM, both the “observable” and “hidden" events are probabilistic and are causal factors to the HMM. This can be written as below formula:

$$P\ (o_i|q_1 \ldots . q_i, q_T\ , o_1, \ldots, o_T)\ = P\ (o_i|q_i)$$

, in which:

- $Q\ =\ q_1\ \ q_2 \ldots .\ \ q_N$ — a set of N states
- $A\ =\ a_{11}\ \ a_{ij} \ldots .\ \ a_{NN}$ — a transition probability matrix A, each a_{ij} indicating the likelihood of transitioning from state i to state j $\sum_{j=1}^{N} \ a_{ij} = 1\ \forall i$
- $O\ =\ o_1\ \ o_2 \ldots .\ \ o_T$ — a string of T observations
- $B\ =\ b_i(o_t)$ — a sequence of observation likelihoods, or emission probabilities, each is the probability of an observation o_T being generated from a state i
- $\pi\ =\ \pi_1, \pi_2, \ldots, \pi_N$ — an initial probability distribution over states, π_i is the probability that the Markov chain will start in state i. Some states j may have $\pi_j\ = 0$, meaning that they cannot be initial states. Additionally, $\sum_{i=1}^{n} \ \pi_i\ = 1$

Table 2. HMM model fit statistic

States	AIC	BIC	Log likelihood
2	57941.69	58072.23	-31489.44
3	55306.96	55533.24	-24858.15
4	53000.64	53340.04	-28952.88
5	53913.75	54383.70	-28133.14
6	56661.43	57279.32	-25331.85
7	**50473.40**	**51256.65**	**-24978.69**
8	55637.18	56603.18	-24978.69
9	54774.65	55940.82	-27313.07
10	51968.22	53351.96	-24882.85

Table 3. HMM transition matrix

	To State 1	To State 2	To State 3	To State 4	To State 5	To State 6	To State 7
From State 1	**0.213**	0.009	0.195	0.045	**0.268**	**0.261**	0.008
From State 2	0.012	**0.731**	0.018	0.050	0.011	0.013	0.166
From State 3	0.194	0.014	**0.209**	0.066	**0.238**	**0.262**	0.016
From State 4	0.116	0.101	0.170	0.184	0.120	0.172	0.137
From State 5	**0.212**	0.007	0.189	0.037	**0.288**	**0.261**	0.006
From State 6	**0.201**	0.008	**0.203**	0.052	**0.254**	**0.272**	0.010
From State 7	0.014	**0.225**	0.029	0.095	0.013	0.023	**0.601**

The purpose of utilizing HMM in this study is to model changes in regulatory phases associated with student affective state. In addition to Markov Chain, which just models probabilistic state changes of observable data, HMM can capture hidden states characterized by changing patterns of emotion that are not readily visible. In light of this, understanding how affective state is associated with regulatory activities can be enhanced by looking at the co-occurrence of this and emotion regulation activities.

For the purpose of determining the most appropriate approach to fitting the HMM model, our affective state data were discretized. The im2pose FER model returns probability values for seven different types of emotion: anger, disgust, fear, happiness, neutral, sadness, and surprise. Using a threshold of 0.5 and higher, the affective state data was discretized into 10 categories based on the dominant emotion per frame.

Based on the calculated AIC, BIC, and log-likelihood statistics of HMM models of 2 - 10 states (Table 2), the seven-state model was selected as it provided the optimal fit for all three criteria. The data were then fit to a seven-state model to determine the best state for each student's emotion per frame/second. A transition matrix that stores the probability of transition between the states is also calculated (Table 3).

4. Results and Findings

4.1. How does Emotional Regulation Occur in Synchronous CSCL?

Table 4 shows the frequency of each type of regulatory interaction, challenges, and adopted emotion regulation strategies.

Table 4. Regulatory activities across phases

Emotion Regulatory Strategies	*f*
Encouragement	2
Increasing Awareness	1
Social Reinforcement	41
Task Structuring	12
Interactions	***f***
Cognitive Interaction	1,466
Socio-emo Interaction	47
Task Execution Interaction	260
Other Interactions	162
Challenges	***f***
Cognitive Challenge	61
Emo-Mo Challenge	22
Social Challenge	91

In total, 66 observable instants of emotion regulation activities were identified in the synchronous CSCL work. Most often, student engaged with social reinforcement (f = 41), follow by task structuring (f = 12). The least adopted strategies are encouragement (f = 2) and following that, increasing awareness (f = 1). These numbers are small in comparison with the overall number of interactions (f_{Total} = 1,935) and regulatory challenges (f_{Total} = 174) in the study. This study's results align with previous studies, which indicated that socially shared emotional regulation during collaborative learning is rare and does not naturally happen in the absence of challenging situations or support to enhance regulation.

4.2. What Types of States can be Identified with HMM based on facial expression recognition?

Our HMM model identified seven latent states, which included various types of regulatory activities are characterized as:

State 1: Segments comprised of two emotions, fear and surprise, and less than 1% of affective state data.

State 2: Segments comprised of 5% affective state data, consisting of 0.23% of anger, 4.7% neutral, and 0.2% surprise.

State 3: Segments comprised more than 50% of affective state data, with neutral accounting for more than 44%, followed by surprise of 5% and less than 0.5% of fear, happiness, and sadness altogether.

State 4: Segment comprised of less than 1% of affective state data, consisting of anger, fear, happiness, and neutral.

State 5: Segment comprised more than 3% of affective state data, consisting of sadness, neutral, disgust, and anger, which account for 2.6%, 0.9%, 0.4%, and 0.3%, respectively.

State 6: Segment comprised of more than 3% of affective state data, consisting of sadness, neutral, disgust, anger, fear, happiness, and surprise.

State 7: Segment comprised around 37% of affective state data, with 16.8% of surprise, 10.9% of neutral, 4.3% of fear, 3.63% of happiness, and less than 1% of anger and disgust respectively.

To summarise, state 1 and 4 mostly did not account for the affective state data, whereas state 3 and 7 account for the majority of the affective state data with less than 10% shared by state 2, 5, and 6.

4.3 How do emotion regulatory activities associated with emotion shifting patter during Synchronous CSCL

To better understand how these hidden states typically occurred, and how they were related to each other, Figure 1 shows the most prominent transition between them. Assignations are marked with arrows, and states, which account for most affective state data, are highlighted in blue.

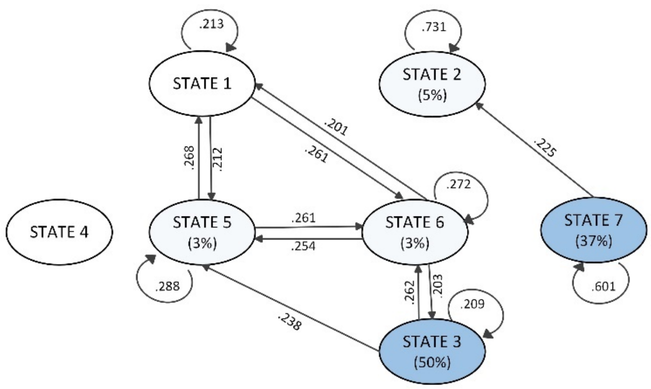

Figure 1. Associations between seven hidden states.

State 7, constituting more than one-third of the data, is characterized by a large proportion of surprise and neutral emotions and, just secondary to the trend of looping back into itself, is most likely to transition into state 2, denoted by dominant neutral emotions.

States 5 and 6, while each only accounting for 3% of affective state data, seem to be the intermediate states encompassing a wide range of emotions. State 1, representing less than 1% of affective state data, includes both fear and surprise equally likely to be transitioned into these two states 5, 6, and vice versa. State 3, which accounts for the greatest proportion of the data and consists predominantly of neutral emotions, is most likely to transition to state 6, followed by state 5.

One interesting observation in this figure is the pattern of loop transition of each state back to itself. This behaviour is evident across all states with the highest probability in state 2 (0.73) and state 7 (0.60).

As part of this analysis, a series of Chi-square tests were conducted to determine the correlation between these latent states and regulatory interactions, challenges, and emotion regulation strategies. The result indicated a statistically significant relationship between the hidden states and different types of interactions X^2 (18, N = 21,080) = 282.97, $p < .001$, challenges X^2 (12, N = 1,676) = 61.75, $p < .001$, and emotion regulation strategies X^2 (18, N = 456) = 57.11, $p < .001$ in collaborative learning. This means that there was a meaningful difference between different learning and regulating phases associated with the affective states changing pattern, captured by FER and HMM. In order to provide a better understanding of the findings, Figure 2 provides a visual representation of the activity types categorized by state according to the percent count.

5. Discussion and Conclusion

The finding from this study makes several contributions to the current literature. First, by examining the changing patterns in the facial emotion recognition data during synchronous CSCL context, this study aimed to reveal the structure of affective states during emotion regulation processes and provide evidence of how AI techniques can be utilized to understand it better. Based on this data, episodes of emotion regulation strategies are marked with relevant transitions of affective states. That is, these latent affective state transitions that can be tracked by using unobtrusive data have the potential to be used as one data channel when investigating measures that are sensitive to the dynamic manifestation of SSRL across phases and times. The findings of this study are important because it not only illuminates the unobservable cognitive processes through its reflection in the changing pattern of facial expression but also demonstrates a methodological and analytical approach to studying video-based data, moving toward unfolding learning regulation, especially emotion regulation in a collaborative context. On that account, the study responded to recent calls for a multidisciplinary effort to advance theoretical understanding and methodological techniques in researching regulated learning (Järvelä et al., 2019; Nguyen et al., 2022).

Second, this study contributes to the growing body of research on the development of synchronous SSRL and collaboration tools, which is particularly important in light of the recent COVID-19 pandemic (Järvelä & Rosé, 2020). Instead of static tools addressing each regulation phase and removing one completely, CSCL tools, especially in synchronous contexts, should adapt and morph based on the changing shapes and needs in students' regulatory processes. Taking this into consideration, the present study revealed hidden affective states and their transition as potentially relevant markers that can be used to measure and assess in-situ changes in students' emotional regulatory development in a timely and efficient manner, making this support possible (Hadwin et al., 2005).

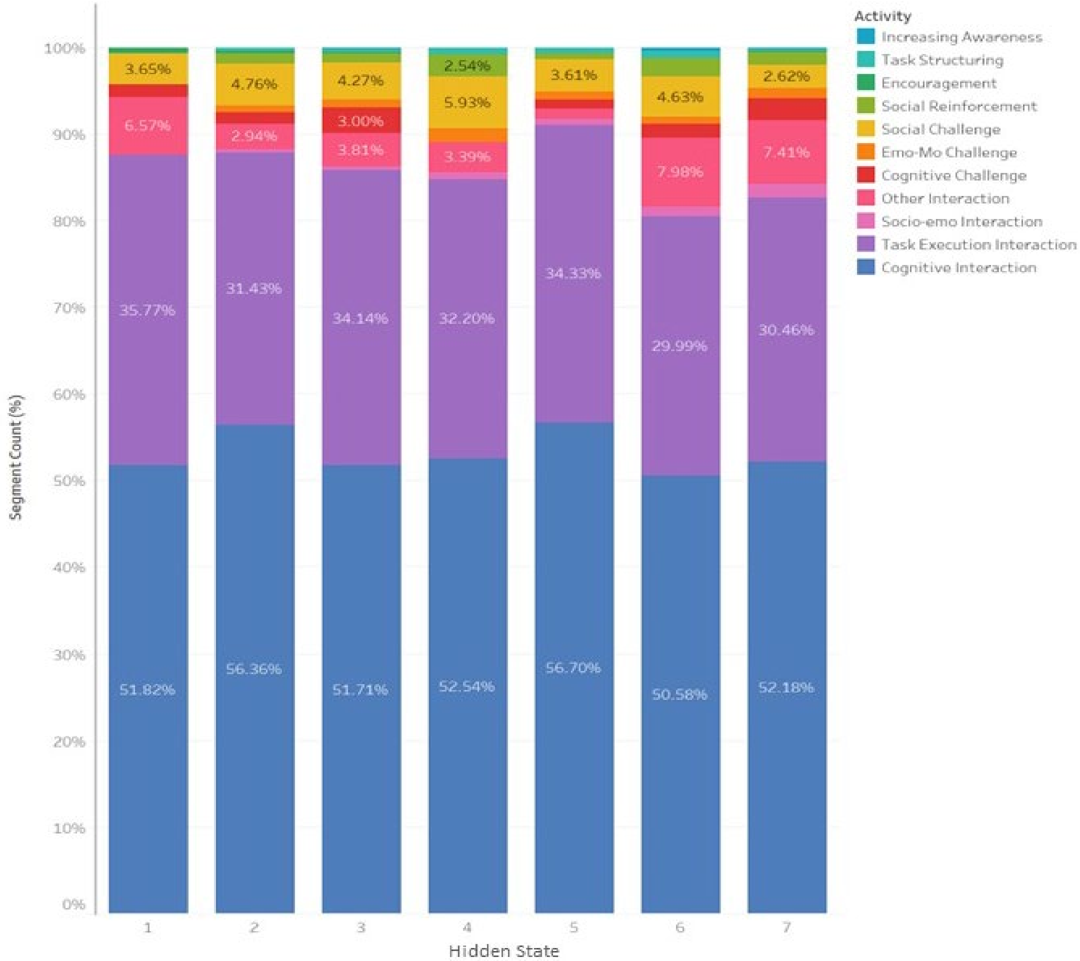

Figure 2. HMM state log counts for regulatory activities across phases

Given the small sample scope and the exploratory nature of utilizing the advanced methodological approach to analyze video data, future efforts are required to ascertain the reliability of our approach. Nevertheless, this also constitutes one of the contributions made by our study. Learning science has begun to see a rise in the use of multimodal and AI-powered analysis tools, but researchers are facing a number of challenges ranging from developing and conducting research design to analyzing and synthesizing complex data. For methodological and theoretical progression, there is a need to share not only successful results but also the overall design of these research and ideas for analysis to harness the potential of unobtrusive data and AI-powered technologies (Harteis et al., 2018). The results of our work not only provide a better understanding of learners' emotion regulation but also establish a foundation for later development of support for predicting and promoting regulation in synchronous online collaborative learning

6. References

Azevedo, R., & Gašević, D. (2019). Analyzing Multimodal Multichannel Data about Self-Regulated Learning with Advanced Learning Technologies: Issues and Challenges. In Computers in Human Behavior (Vol. 96). https://doi.org/10.1016/j.chb.2019.03.025

Coelho, J. P., Pinho, T. M., & Boaventura-Cunha, J. (2019). *Hidden Markov Models : Theory and Implementation Using MATLAB®*. Taylor & Francis Group. http://ebookcentral.proquest.com/lib/rmit/detail.action?docID=5845489

Dindar, M., Jarvela, S., Ahola, S., Huang, X., & Zhao, G. (2020). Leaders and followers identified by emotional mimicry during collaborative learning: A facial expression recognition study on emotional valence. IEEE Transactions on Affective Computing. https://doi.org/10.1109/taffc.2020.3003243

Dindar, M., Järvelä, S., Nguyen, A., Haataja, E., & Çini, A. (2022). Detecting shared physiological arousal events in collaborative problem solving. *Contemporary*

Educational Psychology, *69*, 102050. https://doi.org/10.1016/j.cedpsych.2022.102050

Graham, S. A., Lee, E. E., Jeste, D. v., van Patten, R., Twamley, E. W., Nebeker, C., Yamada, Y., Kim, H. C., & Depp, C. A. (2020). Artificial intelligence approaches to predicting and detecting cognitive decline in older adults: A conceptual review. Psychiatry Research, 284. https://doi.org/10.1016/j.psychres.2019.112732

Hadwin, A. F., Järvelä, S., & Miller, M. (2015). Self-Regulated, Co-Regulated, and Socially Shared Regulation of Learning. In Handbook of Self-Regulation of Learning and Performance. https://doi.org/10.4324/9780203839010.ch5

Hadwin, A. F., Wozney, L., & Pontin, O. (2005). Scaffolding the appropriation of self-regulatory activity: A socio-cultural analysis of changes in teacher-student discourse about a graduate research portfolio. In Instructional Science (Vol. 33, Issues 5–6). https://doi.org/10.1007/s11251-005-1274-7

Harley, J. M., Bouchet, F., Hussain, M. S., Azevedo, R., & Calvo, R. (2015). A multi-componential analysis of emotions during complex learning with an intelligent multi-agent system. Computers in Human Behavior, 48. https://doi.org/10.1016/j.chb.2015.02.013

Harteis, C., Kok, E. M., & Jarodzka, H. (2018). The journey to proficiency: Exploring new objective methodologies to capture the process of learning and professional development. In *Frontline Learning Research* (Vol. 6, Issue 3, pp. 1–5). European Association for Research on Learning and Instruction.

Heerdink, M. W., van Kleef, G. A., Homan, A. C., & Fischer, A. H. (2013). On the social influence of emotions in groups: Interpersonal effects of anger and happiness on conformity versus deviance. Journal of Personality and Social Psychology, 105(2), 262–284. https://doi.org/10.1037/a0033362

Isohätälä, J., Järvenoja, H., & Järvelä, S. (2017). Socially shared regulation of learning and participation in social interaction in collaborative learning. International Journal of Educational Research, 81. https://doi.org/10.1016/j.ijer.2016.10.006

Järvelä, S., & Bannert, M. (2021). Temporal and adaptive processes of regulated learning—What can multimodal data tell? *Learning and Instruction*, *72*, 101268. https://doi.org/10.1016/j.learninstruc.2019.101268

Järvelä, S., & Rosé, C. P. (2020). Advocating for group interaction in the age of COVID19. *International Journal of Computer-Supported Collaborative Learning*, *15*(2), 143–147. https://doi.org/10.1007/s11412-020-09324-4

Järvelä, S., Hadwin, A., Malmberg, J., & Miller, M. (2018). Contemporary Perspectives of Regulated Learning in Collaboration. In F. Fischer, C. E. Hmelo-Silver, S. R. Goldman, & P. Reimann (Eds.), *International Handbook of the Learning Sciences* (1st ed., pp. 127–136). Routledge. https://doi.org/10.4324/9781315617572-13

Järvelä, S., Järvenoja, H., & Malmberg, J. (2019). Capturing the dynamic and cyclical nature of regulation: Methodological Progress in understanding socially shared regulation in learning. *International Journal of Computer-Supported Collaborative Learning*, *14*(4), 425–441. https://doi.org/10.1007/s11412-019-09313-2

Järvenoja, Hanna, Piia Näykki, and Tiina Törmänen. 'Emotional Regulation in Collaborative Learning: When Do Higher Education Students Activate Group Level Regulation in the Face of Challenges?' *Studies in Higher Education* 44, no. 10 (3 October 2019): 1747–57. https://doi.org/10.1080/03075079.2019.1665318.

Khaireddin, Y., & Chen, Z. (2021). Facial emotion recognition: State of the art performance on FER2013. *ArXiv Preprint ArXiv:2105.03588*.

Malmberg, J., Fincham, O., Pijeira-Díaz, H. J., Järvelä, S., & Gašević, D. (2021). Revealing the hidden structure of physiological states during metacognitive monitoring in collaborative learning. *Journal of Computer Assisted Learning*, *37*(3), 861–874.

Malmberg, J., Järvelä, S., & Järvenoja, H. (2017). Capturing temporal and sequential patterns of self-, co-, and socially shared regulation in the context of collaborative learning. Contemporary Educational Psychology, 49. https://doi.org/10.1016/j.cedpsych.2017.01.009

Nguyen, A., Järvelä, S., Wang, Y., & Róse, C. (2022). Exploring Socially Shared Regulation with an AI Deep Learning Approach Using Multimodal Data. *Proceedings of International Conferences of Learning Sciences (ICLS)*. https://2022.isls.org/proceedings/

Rogat, T. K., & Adams-Wiggins, K. R. (2015). Interrelation between regulatory and socioemotional processes within collaborative groups characterized by facilitative and directive other-regulation. Computers in Human Behavior, 52. https://doi.org/10.1016/j.chb.2015.01.026

Tyng, C. M., Amin, H. U., Saad, M. N. M., & Malik, A. S. (2017). The influences of emotion on learning and memory. In Frontiers in Psychology (Vol. 8, Issue AUG). https://doi.org/10.3389/fpsyg.2017.01454

Xiao, Y., & Yang, M. (2019). Formative assessment and self-regulated learning: How formative assessment supports students' self-regulation in English language learning. *System*, *81*, 39–49.

Zhang, J., Yin, Z., Chen, P., & Nichele, S. (2020). Emotion recognition using multimodal data and machine learning techniques: A tutorial and review. Information Fusion, 59. https://doi.org/10.1016/j.inffus.2020.01.011

Zhu, X., Lei, Z., Liu, X., Shi, H., & Li, S. Z. (2016). Face alignment across large poses: A 3d solution. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 146–155.

Zimmerman, B. J. (2013). From Cognitive Modeling to Self-Regulation: A Social Cognitive Career Path. Educational Psychologist, 48(3). https://doi.org/10.1080/00461520.2013.794676