

RESEARCH ARTICLE

Open Access



Unveiling emotion dynamics in problem-solving: a comprehensive analysis with an intelligent tutoring system using facial expressions and electrodermal activities

Juan Zheng^{1*} , Shan Li^{1,2} , Tingting Wang³ and Susanne P. Lajoie³

*Correspondence:

Juan Zheng

juz322@lehigh.edu

¹Department of Education and Human Services, Lehigh University, Bethlehem, PA 18015, USA

²Department of Community & Population Health, Lehigh University, Bethlehem, PA 18015, USA

³Department of Educational & Counselling Psychology, McGill University, Montreal, QC, Canada

Abstract

Emotions play a crucial role in the learning process, yet there is a scarcity of studies examining emotion dynamics in problem-solving with fine-grained data and advanced tools. This study addresses this gap by investigating the emotional trajectories during self-regulated learning (SRL) phases (i.e., forethought, performance, and self-reflection) among 47 medical students utilizing an intelligent tutoring system. Real-time facial expressions were analyzed through recurrence quantification analysis alongside an examination of electrodermal activities (EDA) across the SRL phases. The findings reveal that emotion stability varied across SRL phases, with students exhibiting more stable emotions during the performance phase. Compared to the forethought and self-reflection phases, students had less frequent and lower intensity of emotional arousal in the performance phase. Moreover, we found that students with better performance demonstrated more stable emotions in the forethought phase, less stable emotions in the self-reflection phase, and a higher level of emotional arousal in the self-reflection phase. These insights highlight the temporal and dynamic nature of emotions in SRL, offering methodological and educational implications for leveraging facial expressions and EDA to monitor and enhance students' emotional experience during problem-solving.

Introduction

Emotions are dynamic in nature. Emotion states, as described by Ahmed et al. (2013), are the transient and momentary emotional experiences occurring within a single episode of academic life. These emotion states are not static but rather continuously evolve in response to various situational cues and contextual factors (Goetz et al., 2015). This means that emotions are experienced within specific episodes or situations, and their intensity and nature can vary depending on the context in which they occur. Situational cues, including environmental stimuli, social interactions, and individual perceptions of the learning context, play a pivotal role in influencing the fluctuation of emotion state

In the process of clinical problem-solving, emotions fluctuate in response to ever-changing learning processes (Zheng, Li et al., 2023). For example, consider a student who attempts to independently solve a challenging clinical problem. At the beginning, they might feel happy and confident as they engage with the problem. However, when encountering difficult concepts or struggling to understand certain topics, their emotions may shift to frustration or anxiety. With perseverance and effective learning strategies, they can eventually overcome these challenges, leading to a sense of pride and happiness. Throughout these problem-solving processes, student emotions evolve and fluctuate in response to the dynamic nature of their learning experience. In other cases, students with similar levels of emotions may demonstrate significant variations in emotion fluctuation during a task (Reitsema et al., 2022). For example, one student may constantly experience medium-level anxiety throughout the process of problem-solving, whereas the other student may experience both high and low levels of anxiety depending on his/her progress on solving the problem. All these fluctuations and patterns in the intensity, duration, and quality of emotional experiences over time are referred to as emotion dynamics (Zheng, Li et al., 2023). Emotion dynamics can be quantified by measures of variability, instability, or inertia of emotions (Houben et al., 2015). A variety of studies have examined the emotion dynamics using self-report questionnaires (Niculescu et al., 2016; Peterson et al., 2015). However, self-report data fails to capture emotions at a fine-grained size. Current measurements of emotions are insufficient to assess the temporal structures of emotions as they naturally occur in real-time (Jenkins et al., 2020). Therefore, there is a growing call to assess emotion dynamics with fine-grained data and more advanced tools.

Increasingly, researchers are extracting indicators of emotions from trace data, including facial expressions and electrodermal activity (EDA). Facial expressions can be automatically recorded and analyzed to identify learners' real-time emotional states (Lajoie, Zheng et al., 2021; Taub et al., 2020). EDA is one of the popular physiological measures that capture learners' electrical changes at the surface of the skin. Unlike self-report questionnaires, EDA provides arousal information that is challenging for learners to mask (Dan-Glauser & Gross, 2013). Changes in arousal, valence, and activation can be associated with emotions. More importantly, researchers can simultaneously and unobtrusively collect facial expressions and EDA without interrupting students' learning processes. These two methods provide time-series data, allowing for fine-grained analysis of emotion dynamics.

A few studies have attempted to explore the relationships between emotions and learning using either facial expressions or EDA. For instance, Buono et al. (2020) used automated facial recognition software (iMotions, 2018) to identify anger, fear, frustration, confusion, happiness, and surprise. They found that frustration was negatively related to the planning strategy, and happiness was positively associated with the monitoring strategy. Similarly, taking advantage of an emotion recognition software to identify emotions from facial expressions, Taub et al. (2020) found that happiness immediately after relevant activities and confusion after positive feedback positively predicted students' performance. More recently, Li et al. (2021) explored students' emotion variability using facial expressions and found that emotion variability contributed to students' performance. In addition to facial expressions, EDA was found to be positively related to self-reported workload and cognitive regulation (Dindar et al., 2020). More

interestingly, EDA parameters changed according to the challenges of learning (Haataja et al., 2016). However, the joint role of EDA parameters and the emotions detected from facial expressions in clinical problem-solving is still underexplored.

The purpose of this study is to contribute to the literature on emotion dynamics by: (1) collecting the time-series data of students' facial expressions and EDA in real time as students independently solve an authentic problem in an intelligent tutoring system, (2) examining the arousal level of emotions across the problem-solving process with the real-time EDA data, and (3) analyzing fine-grained emotions with an advanced analytical technique, i.e., recurrence quantification analysis (RQA). Specifically, RQA is a non-linear method that quantifies the dynamics of temporal sequences of change over time by detecting 'recurrent events' in a time series (Dale et al., 2011; Jenkins et al., 2020). RQA generates a few metrics to describe the temporal structures and patterns of emotional change (Main et al., 2016). Another unique advantage of RQA is that it requires no *priori* assumptions about data in terms of normality, independence, equal variance, and stability (Jenkins et al., 2020). To our knowledge, this study is the first to use RQA to explore the temporal structures of students' emotions in the context of clinical problem-solving.

Theoretical framework and literature review

Clinical problem-solving in an intelligent tutoring system

Clinical problem-solving is a fundamental aspect of medical education and practice due to its critical role in training competent healthcare professionals. Clinical problem-solving refers to a dynamic and recursive thinking process wherein medical students and practitioners collect evidence, propose hypotheses, plan or implement treatment, evaluate outcomes, and reflect on the clinical reasoning process (Kuiper, 2013). Research studies showed that expert medical practitioners demonstrated more efficient problem-solving strategies (Benner et al., 2008), whereas novice medical students faced the challenge of striking a balance between efficiency and accuracy when making decisions under uncertainty (Isen et al., 1991). Solving clinical problems effectively requires medical students to not only leverage their medical knowledge and cognitive reasoning skills, but also actively engage in maintaining, monitoring, and assessing their own thinking processes (Wang et al., 2023). The demanding and high-risk nature of clinical problems necessitates additional intervention or support for medical students to thrive in medical school. Educational interventions aimed at enhancing clinical problem-solving have investigated the integration of technology-enhanced learning platforms, such as intelligent tutoring systems, as a potential approach.

Intelligent tutoring systems offer medical students extensive guided practice at a cost-effective rate while motivating them through diverse media and materials (Dharmathilaka et al., 2023). For example, Suebnukarn and Haddawy (2006) developed an intelligent tutoring system named COMET to stimulate group discussion and recommend peer assistance. Kazi et al. (2013) enhanced the capabilities of intelligent tutoring systems for clinical problems by incorporating a comprehensive medical knowledge source. In another study, Dharmathilaka (2023) developed an intelligent tutoring system allowing students to collect the patient's medical history, perform physical examinations, request laboratory tests, and ultimately make informed diagnostic and treatment decisions. These reviewed intelligent tutoring systems showcase how they can support medical students in solving clinical problems. As pointed out by Elstein and Schwarz (1978),

intelligent tutoring systems designed for clinical reasoning should provide three essential functions: hypothesis testing, pattern matching, and categorization. The intelligent tutoring system in this study was developed with consideration of all these supports, which will be explained in detail. Moreover, the system was designed in accordance with self-regulated learning theory to provide scaffolding for medical students in monitoring and evaluating their thinking processes. As students regulate themselves to resolve knowledge conflicts and navigate feelings of success or failure, they may encounter a range of intense emotions. Therefore, the following section will explore the potential role of emotions in self-regulated learning.

Emotions in self-regulated learning

Self-regulated learning (SRL) theories describe how learners strategically self-regulate their cognitive, metacognitive, affective, and motivational aspects of learning to achieve personal goals (Zimmerman, 1990). In this sense, emotion is an integral component of SRL. In fact, researchers have reached a consensus that emotions play an important role in SRL. According to Pekrun's (2006) control-value theory, positive activating emotions (e.g., happy) promote the use of flexible, deep strategies, whereas negative activating emotions (e.g., anxiety) facilitate rigid ways to solve problems. Muis et al. (2018) claimed that emotions, such as curiosity and surprise, could influence how students plan, set goals, and use cognitive and metacognitive strategies. To step further, in the "Metacognitive and Affective model of Self-Regulated Learning" (MASRL) model, Efklides (2011) describes how state-like emotions could be triggered in different SRL phases, i.e., *forethought*, *performance*, and *self-reflection* (Zimmerman, 2000). In the *forethought* phase, students prepare for efforts by analyzing the task. In this phase, students either experience automatic processing on familiar tasks or undergo effortful processing due to the mismatch between task features and prior knowledge. In the case of automatic processing, students are likely to have neutral or moderately positive emotions without increased physiological activity (Carver & Scheier, 1998). In the case of effortful processing, students may experience increased emotional arousal. In the *performance* phase of SRL where students execute the task by employing learning strategies, the fluency of information processing and the rate of progress determine the intensity and the type of emotions students may experience (Ainley et al., 2005). Particularly, students may have increased physiological activities due to interruption and conflict in cognitive processing. Finally, the outcome of cognitive processing and external feedback trigger emotions in the *self-reflection* phase. Students may experience either positive or negative emotions depending on the learning outcome and whether the learning outcome is expected or not. To sum it up, control-value theory reveals how emotions influence SRL strategies, whereas MASRL model demonstrates how the type and intensity of emotions change across the three phases of SRL. These prevailing theoretical models highlight the dynamic nature of emotions throughout SRL phases, thereby providing a foundation for investigating emotion dynamics within SRL through the utilization of fine-grained measures.

Measuring emotions in real time with facial expressions and EDA

Emotion dynamics reflect the short-term fluctuations of moment-to-moment emotions. Measuring emotion dynamics directly is not feasible; instead, it requires quantifying a

set of temporal features of emotions. Advanced measurement techniques capable of capturing emotions in real time at a fine-grained level are necessary to accurately record these temporal features. Facial expressions and EDA are two of these kinds, which have been widely used in a range of empirical studies. For instance, Taub et al. (2020) used students' facial expressions to investigate how emotions associated with the key actions of problem-solving in SRL. They found that students were happy when reading relevant contents, but they were confused when conducting laboratory tests. As another example, Lajoie et al. (2021) measured emotions based on students' facial expressions. They examined the co-occurrences of emotions and SRL processes and found that angry and surprise was more likely to occur in the *performance* and *self-reflection* phases of SRL. In terms of EDA, Shukla et al. (2019) identified 40 EDA features in a review of 25 studies. They found that the same EDA features had been used for both arousal recognition and valence recognition. They also found that subject-dependent EDA analysis was better than subject-independent analysis, indicating EDA analysis should be closely associated with the individual and the research context. Moreover, Shukla et al. (2019) found that skin conductance responses (i.e., SCRs) are the most used features to indicate emotional arousal. SCRs are event-related features that reflect participants' short-term responses to stimuli. The number of significant SCRs and the amplitude of SCR can be used to indicate the arousal level of emotions. Harley and his colleagues (2015) synchronized facial expressions, EDA, and self-reported emotions as students learned a complex science topic. They found a high agreement between facial expressions and self-report data. However, EDA provided information that differed from that of facial expressions and self-report emotions. Facial expressions and EDA may reflect different aspects of emotions, suggesting that both should be included in analyses to provide a comprehensive image of emotions.

Recurrence quantification analysis

Recurrence Quantification Analysis (RQA) is an advanced method that generates multiple metrics from time-series data to quantify the structure of emotion dynamics (Richardson et al., 2014). Moreover, RQA can also be used to assess emotion dynamics with categorical time-series data (Lichtwarck-Aschoff et al., 2012). Of particular relevance to the current study are the metrics of percent recurrence (%REC) and laminarity (%LAM). As displayed in Table 1, %REC measures how often a person experiences the same emotion over time (i.e., the degree to which the same emotional state reoccurs over time). %LAM captures the degree same emotions occur in repeating sequences (i.e., the ratio

Table 1 The selected RQA measures to quantifying the complexity of emotions

Variable	Definition	Meaning
Percent recurrence (%REC)	Percentage of recurrence points in a recurrence plot (RP). %REC = Sum of recurrent points in the RP / size of RP. %REC = Sum of recurrences / N (N - 1), where N refers to the length of the behavioral sequence.	How often does an individual experience the same emotion over time (i.e., the degree to which the same emotional state occurs over time)?
Laminarity (%LAM)	The proportion of recurrent points forming vertical line structures. %LAM = Sum of vertically adjacent recurrent points/sum of recurrent points.	To what extent do the same emotions occur in repeating sequences? (i.e., the ratio of repeating sequences of the same emotion to recurring emotions).

(Marwan et al., 2002; Meinecke et al., 2020; Wallot, 2017)

of repeating sequences of the same emotion to recurring emotions). The calculation of these RQA metrics is based on the distribution of the recurrent points in a recurrence plot (RP) (see Fig. 1). The RP visualizes the recurrence of emotions by plotting a discrete time-series of emotion states on both the *x* and *y*-axis of a two-dimensional grid. As displayed in Fig. 1, the three adjacent points that form a vertical or horizontal line represent a repeating sequence of the same emotion, e.g., Surprise (SU) → Surprise (SU) → Surprise (SU). %REC equals the percentage of recurrent points in a RP, whereas %LAM is the proportion of recurrent points forming vertical line structures. In the example of emotion sequence (Fig. 1), surprise reoccurs 30 times, sad 6 times, anxious 2 times, angry 2 times, and happy 0 time. $\%REC = (30+6+2+2+0)/14*13 = 21.98\%$. %LAM = the number of adjacent black dots / the number of black dots = 32/40 = 80%.

Several studies have successfully used recurrence quantification analysis (RQA) to examine the temporal structures of emotions. For instance, Main et al. (2016) applied RQA on emotions coded from conflict discussions between mother and children, assessing emotion synchrony between mother-child dyads. The study of Jenkins et al. (2020) was the first to utilize RQA for determining the emotion variation over time. They measured students' daily emotions consecutively for two weeks and found that RQA metrics could better predict emotion dynamics than traditional statistics (i.e., mean and standardization). While prior studies have successfully utilized RQA to investigate emotions, there remains a gap in the literature regarding the exploration of students' emotion dynamics using RQA on facial expressions collected during the process of solving clinical problems within an intelligent tutoring system.

Current study

This study aims at examining emotion dynamics in the context of clinical problem-solving. Clinical problem-solving is a complex thinking and reasoning process where

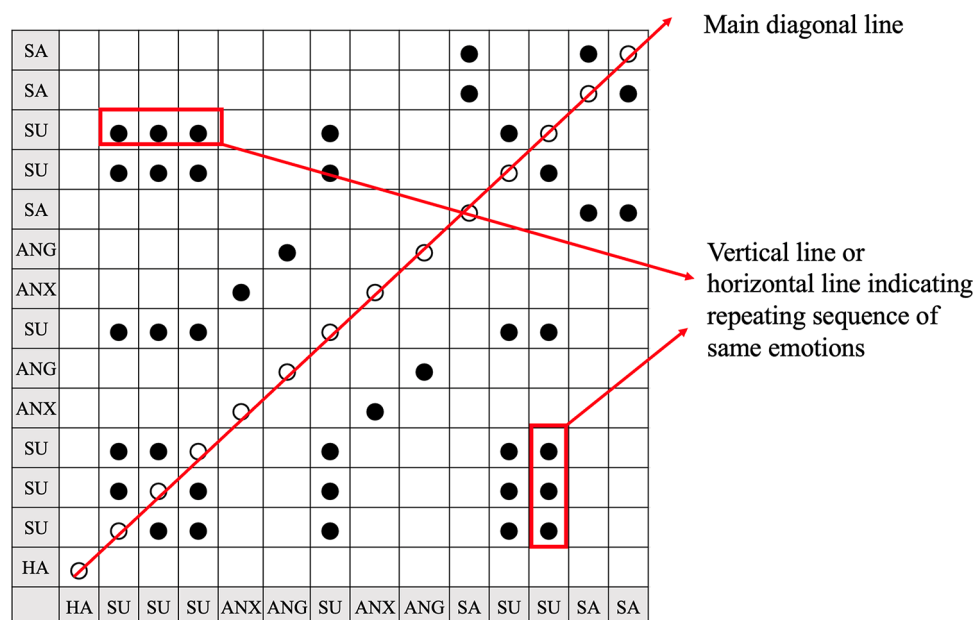


Fig. 1 An example of a recurrence plot. Note HA=Happy, SU= Surprise, ANX= Anxiety, ANG= Angry, SA= Sad. The emotion sequence is plotted on both the *x* and *y*-axis. The black and white dots are placed in positions where the same behavior within the sequence reoccurs. The white dots form the main diagonal line, and the recurrence plot is symmetrical about its main diagonal line. The main diagonal line is excluded when calculating the RQA measures

medical students regulate themselves to monitor and control their cognitive and emotional processes while making correct or incorrect diagnoses of patients (Eva, 2005). In clinical problem-solving, students' state-like emotions dynamically fluctuate alongside their phases of self-regulated learning (SRL). Identifying students' emotion dynamics in clinical problem-solving will enable intelligent tutoring systems to scaffold learners in real time by providing correspondent emotional interventions. Therefore, we measure emotion dynamics by examining both the temporal patterns of emotion sequences and the varying levels of emotional arousal. More specifically, we quantify the temporal structures of emotion sequence with the RQA metrics of %REC and %LAM. We use EDA indices (i.e., the number and amplitude of SCRs) to indicate the arousal levels of emotions. This study addresses the following research questions: (1) Do students demonstrate different temporal structures of emotions across the phases of SRL (i.e., forethought, performance, and self-reflection)? (2) Do students have different arousal levels of emotions across the SRL phases? and (3) What is the relative importance of the temporal structure of emotions and the arousal levels of emotions in specific SRL phases in predicting students' performance? To our knowledge, this is the first study in the literature that examines emotion dynamics in clinical problem-solving. Based on the findings of prior research (Zheng, Lajoie et al., 2023) and the assumptions of SRL and the control-value theories (Pekrun, 2006), we hypothesize that students will have a higher level of emotional arousal during the performance and self-reflection phases of SRL than in the forethought phase. Moreover, the emotional arousal and temporal structures of emotions during the self-reflection phase may be the most important predictors of students' performance among the three SRL phases.

Methods

Participants

This study was part of a large project aiming to help medical students practice clinical problem-solving skills in an intelligent tutoring system. Participants in this study were 47 medical students from a large North American university who were asked to diagnose a virtual patient case in an intelligent tutoring system. Six participants did not consent to report their demographic information. The remaining sample consisted of 24 female students (56.36%) and 17 male students (43.63%). The age of the participants ranged from 20 to 33, with an average age of 24.10 ($SD=3.51$). Among these participants, 51% and 24% self-identified them as Caucasian and Asian, respectively. The ethnicities of the other participants were either Arab, Jewish, or Mixed. The majority of students were in their second year of medical school ($n=30$, 73.17%). There were four students in their fourth year, three students in their third year, and four students in their first year of medical school. They all had no experience in diagnosing virtual patients. But all participants had completed a prerequisite 7-week course on endocrinology, metabolism, and nutrition.

Learning environment and task

The BioWorld system is an intelligent tutoring system designed for medical students to practice clinical problem-solving skills (Lajoie, Li et al., 2021). The system provides students with the necessary cognitive and metacognitive tools to solve virtual patient cases in SRL (Fig. 2). Specifically, students accomplish the task by analyzing the patient (i.e.,

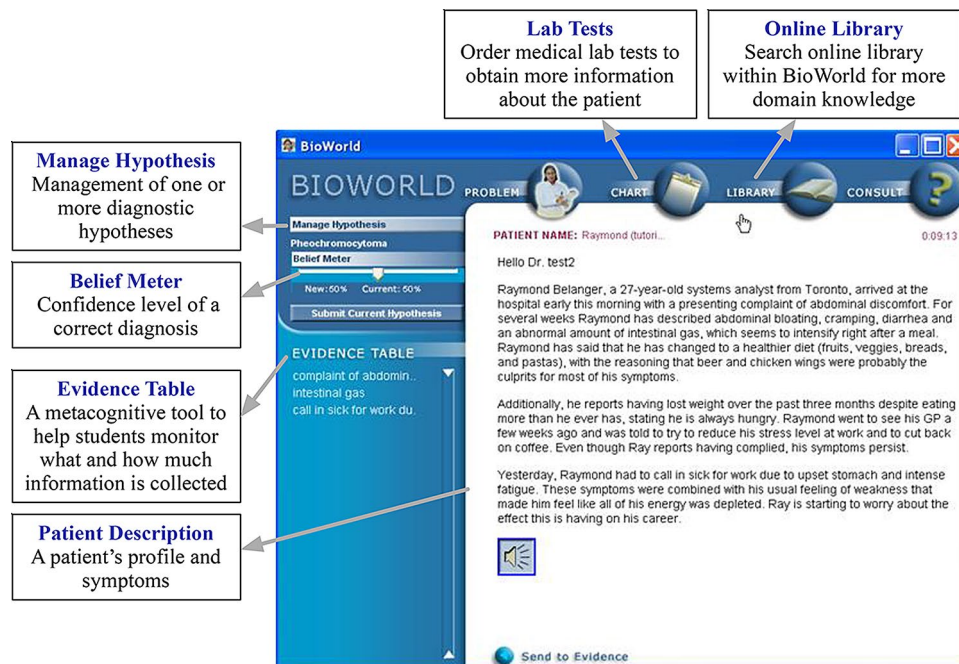


Fig. 2 The main interface of the system

forethought), performing the diagnosis (i.e., performance), and reflecting on their problem-solving processes and outcomes (i.e., self-reflection). Students begin solving a clinical problem by reading a patient's history, which includes the symptoms and medical records that a doctor usually obtains from a patient. At this phase, students familiarize themselves with the task by highlighting relevant evidence, which will be automatically saved in the evidence table for later reference. For example, a student may collect symptoms such as diarrhea, weight loss, as important evidence items from reading the patient history. In the performance phase, students propose initial hypotheses (i.e., diagnose) and run medical lab tests to confirm or disconfirm their hypotheses. For instance, students may conduct the anti-endomysia antibody test within the system. If the test result is positive, students will likely diagnose the patient with a Celiac disease. In addition to lab tests, an embedded medical library is available for students to search for unfamiliar terminology or related knowledge about the case. Students may organize relevant evidence and lab test results in favor of, or against, a specific hypothesis, during which a final hypothesis will be reached. After submitting a final diagnosis, students cannot go backward to change their submission. Instead, students will reflect on their diagnosis by reorganizing relevant evidence items and writing a summary to reproduce the reasoning process. Students extend reflection processes to the reading of individualized feedback where students compare their solutions with an aggregated expert solution. The activities in the reflection phase are useful for another loop of self-regulated learning in the next task.

Participants in this study were asked to diagnose a patient who had Diabetes Mellitus Type 1. Diabetes Mellitus Type 1 is most diagnosed in adolescents affecting millions of people worldwide (Menart-Houtermans et al., 2014). A panel of experts, including medical professionals and learning scientists, created the case and the case solution based on an authentic patient. When students submit a final diagnosis, the system will provide

them with individualized feedback regarding the evidence items collected by experts, the lab tests ordered by experts, and a written case summary of experts. Students' performance is calculated as the evidence/lab tests match with experts' solutions.

Measures

Facial expressions

Participants' facial expressions were recorded by a camera which was mounted above the monitor of the computer participants were using. Facial expression videos were automatically saved as WMV files with a start time and a resolution of 1600*1200. The video data were then analyzed with the FaceReader software to classify six basic emotions (i.e., happy, sad, angry, surprised, disgusted, and scared) and neutral states, following Ekman and Friesen's (1978) emotion action coding. FaceReader uses the Active Appearance Model to model the location of 500 key points in the face and the facial texture of the area entangled by these points. Studies have validated the accuracy of FaceReader. For instance, Loijens et al. (2015) found high consistency between FaceReader performance and manual coding of basic emotions (84.8 – 95.9%). Similarly, Lewinski et al. (2014) reported that FaceReader 6.0 recognized 88% of emotional labels in two public datasets (i.e., Emotional Facial Expression Pictures and Amsterdam Dynamic Facial Expression Set), nearing the 85% accuracy of human coding. These findings support the reliability of FaceReader in automatically detecting emotions. The outputs of FaceReader include the dominant emotional states (a record is added if the emotion is active for at least 0.5 s) and the timestamp information about the onset and transition of each emotional state. The distribution of the six basic emotions across the three phases of SRL is displayed in Table 2. Moreover, FaceReader has been validated with human coders by prior researchers (Chentsova-Dutton & Tsai, 2010; Harley et al., 2015). It provides an unobtrusive way to measure real-time emotional states when students learn to solve a patient case.

Electrodermal activities

Electrodermal activity (EDA) is used to define the continuous variation in the electrical characteristics of the skin, including skin conductance, galvanic skin response (GSR), electrodermal response (EDR), sympathetic skin response (SSR), and psychogalvanic reflex (PGR) (Braithwaite et al., 2013). In particular, skin conductance is most widely used to quantify the electrical reaction of the skin. Skin conductance can be distinguished as tonic skin conductance (skin conductance level, SCL) and phasic skin conductance (skin conductance responses, SCRs). Tonic skin conductance changes relatively slowly without any discrete environmental event or external stimuli, whereas phasic skin conductance (i.e., peaks, SCRs) changes rapidly in accordance with short-term events and discrete environmental stimuli (Braithwaite et al., 2013). The SCR onset, SCR

Table 2 The distribution of basic emotions across the phases of SRL

Basic emotions	Sum of emotions		
	Forethought	Performance	Self-reflection
Sad	123	285	196
Disgusted	68	660	350
Happy	281	324	179
Angry	123	1447	951
Surprised	41	518	372
Scared	51	188	55

rise time, SCR recovery time, SCR frequency, and SCR amplitude are well-recognized EDA features (Shukla et al., 2019). SCR onset is the timepoint where skin conductance level first rises above the pre-set threshold (0.05 μ S). SCR amplitude is calculated from the onset of SCR to the peak value within the SCR. SCR frequency refers to the number of SCRs. SCR rise time is the time taken from SCR onset to reach peak amplitude, whereas SCR recovery time is the time between peak and recovery point (half of the amplitude). SCR duration refers to the sum of onset time and recovery time.

We used two different devices (i.e., Biopac and Q-sensor) to measure students' EDA, simply based on the accessibility of the devices for data collection at the time. Biopac (Braithwaite et al., 2013) produces exosomatic measures of skin conductance with a configuration of 200 samples per second (sampling rate=200). The EDA data of fifteen participants were collected by Biopac in this study. Q-sensor is a bracelet convenient to configure to measure EDA signals (sampling rate=32) (Kapoor et al., 2007). In this study, the EDA data of fifteen participants were collected by Q-sensor. Biopac placed traditional electrodes on the thenar and hypothenar eminences of the palm to record skin conductance response, whereas Q-sensor chose the inner side of the left wrist as a less intrusive recording location (Tomko, 2015). Researchers found traditional palmar measurements were significantly correlated with wrist measurements (van Dooren et al., 2012). Tomko (2015) compared the correlation between Q-sensor and Biopac and found that the correlations between these two devices range from medium to large. In particular, correlations between Q-sensor and Biopac SCRs are larger than that of SCL. Thus, relying on SCRs for EDA analysis improves the consistency between Biopac and Q-sensor results.

Furthermore, we used Neurokit to down sample the raw Biopac EDA data to an equivalent sampling rate of 32 Hz. Neurokit was then used to identify SCRs from raw signals (unfiltered SCL) collected from both devices. Neurokit is an open-source, community-driven, and user-centered package for physiological signal processing (Makowski et al., 2021). The source code of Neurokit is available under the MIT license on GitHub (<https://github.com/neuropsychology/NeuroKit>). It provides a high-level integrative function to process multiple bio-signals. In terms of the EDA processing technique, Neurokit is different from other pure EDA signal processing toolboxes. For example, LedaLab provides researchers the option to choose either *continuous decomposition analysis* that decomposes EDA data into continuous tonic and phasic activity or *discrete decomposition analysis* that decomposes EDA data into a tonic component and discrete phasic components (Benedek & Kaerbach, 2010a, 2010b). Nevertheless, Neurokit uses the methods of convex optimization to identify SCRs based on Bayesian statistics, mathematical convex optimization and sparsity (Greco et al., 2016). More specifically, computes the number, onset, and amplitude of SCRs after filtering the noise of a signal (Greco et al., 2016). The frequency and amplitude of significant SCRs are critical parameters to assess participants' level of emotional arousal (Horvers et al., 2021). For this study, we calculated the SCR frequency and amplitude for each student at each phase of SRL.

Emotion and SRL alignment

FaceReader outputs of emotion states and EDA signals were manually aligned with the system log files so that we could identify students' emotional responses at each phase

of SRL. We divided the clinical problem-solving process into the three phases of SRL based on three key time points recorded in the system log files. For the first time point, students have read the patient description and are about to solve the task (Time 1). Regarding the second time point, students submit their final diagnosis (Time 2). They finish reading their feedback and are about to exit from the task at the third time point (Time 3). More specifically, the *forethought* phase starts from the time students load the page and ends at Time 1. The *performance* phase starts from Time 1 and ends at Time 2, whereas the *self-reflection* phase starts at Time 2 and ends at Time 3. The outputs of FaceReader and the raw EDA signals were divided into three parts based on the three timestamps.

Procedures

Participants who had signed the consent form were asked to complete a pre-survey on Survey Monkey to report their demographic information. Research assistants then gave participants step-by-step instructions to ensure that every student is fully informed about data collection settings. Afterward, a training session was provided to participants on how to use the system to solve clinical problems. Participants practiced on a sample case (i.e., diagnose Celiac disease) to get familiar with the learning environment. After that, participants started to complete the task (i.e., Diabetes Mellitus), during which their facial expressions, EDA data, and log files were automatically saved in the server. During the entire process, research assistants were available to provide technical support. The data analyzed in this study was collected as part of a large and ongoing project. Only the measures relevant to the analyses in this study were discussed above. The facial expression data of four participants was missing due to technical malfunctions (8.5% of the sample), resulting in a sample of 43 for analyzing facial expressions to address the first research question. The EDA data of seventeen participants (34% of the sample) were lost because EDA devices did not capture the entire clinical problem-solving process, leaving a sample of 30 for EDA analysis to answer the second research question. Consequently, a total of 29 participants had both facial expressions and EDA data available for addressing the third research question. To ensure data quality and coherence, we used the same software (i.e., FaceReader) and algorithms (i.e., NeuroKit) to analyze the data of participants.

Data analysis

To answer the first research question, we used the RQA to analyze the temporal structures of emotions among the three phases of SRL. Specifically, the R package of 'crqa' was used to perform the RQA on students' emotional states (Coco & Dale, 2014). We then used Friedman tests to compare the differences in the RQA measures (i.e., %REC and %LAM) among the three phases of SRL, considering that the normality assumption was violated.

To address our second research question, we used Neurokit to preprocess the EDA data in the Scientific Python Development Environment (Raubaut, 2018). Particularly, we used the Makowski algorithm within Neurokit to filter out the noise of EDA raw data and identify the tonic and phasic levels of skin conductance (Makowski, 2020). The outputs include the filtered EDA signal, the phasic and tonic components, SCR onsets, SCR peaks, SCR duration, and SCR amplitudes (see Fig. 3). As discussed above, SCR

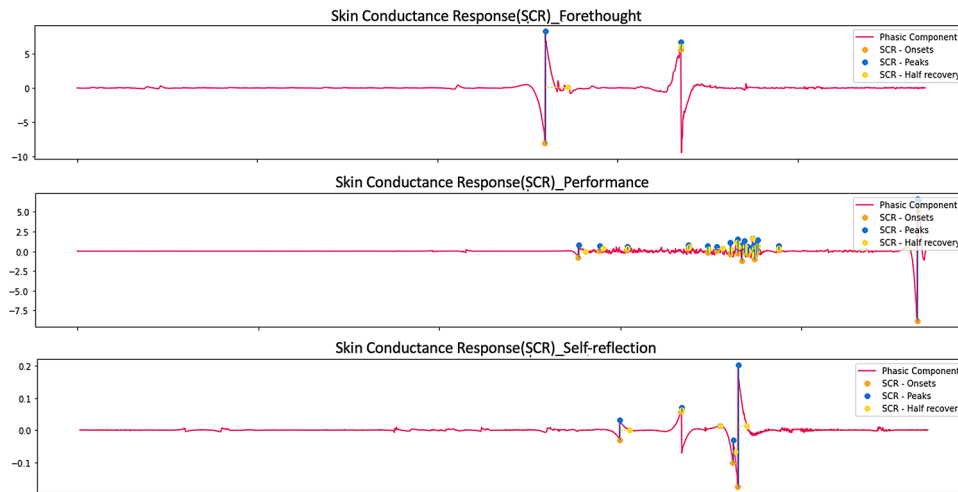


Fig. 3 An example of processed EDA signals

duration, the number, and amplitude of SCR peaks are commonly used as indicators of students' emotional arousal levels. We then compare the differences in SCR numbers per minute, SCR duration, and SCR amplitude across the three phases of SRL using the Friedman tests.

Finally, the assessment method of Averaging over Orderings proposed by Lindeman, Merenda, and Gold (LMG) (Lideman et al., 1980) was applied to answer the third research question. Specifically, the LMG is a dominant method to assess the relative importance of independent variables to a dependent variable in multiple linear regression, especially when the independent variables are correlated (Grömping, 2006). In simple linear models where all regressors are uncorrelated, each regressor's contribution to the model is the R^2 from univariate regression. To address the issue of correlations among regressors, Lideman et al. (1980) proposed averaging sequential sums of squares over all ordering of regressors since the sequence the regressors are added in a multiple linear regression model influences the sum of squares. Therefore, LMG method takes account of the variable orderings when examining the proportion of variance explained by predictors (Lideman et al., 1980). LMG is one of the most computer-intensive methods that require a sample size bigger than the number of predicting variables (Bi, 2012; Grömping, 2006). The formulae denoted as LMG can be written as:

$$LMG(x_k) = \frac{1}{p!} \sum_{r\text{permutation}} seqR^2(\{x_k\} | r) \quad (1)$$

$$seqR^2(\{x_k\} | S_k(r)) = R^2(\{x_k\} \cup S_k(r)) - R^2(S_k(r)) \quad (2)$$

$S_k(r)$ represents the set of predictors entered into the linear model before the predictor x_k in the order r , while $seqR^2(\{x_k\} | S_k(r))$ represents the portion of R^2 allocated to predictor x_k in the order r . Take this study as an example, we compare the relative importance of twelve regressors to a dependent variable, i.e., diagnostic performance. The twelve regressors include six variables related to the temporal structures of emotion sequence (i.e., % REC – forethought, % REC – performance, % REC – reflection, % LAM – forethought, % LAM – performance, % LAM – reflection) and the other six variables concerning to the arousal levels of emotions (i.e., SCR number – forethought, SCR

number – *performance*, SCR number – *reflection*, SCR amplitude – *forethought*, SCR amplitude – *performance*, SCR amplitude – *reflection*). For the twelve regressors ($p=12$), there are $12!$ ($12 \times 11 \times 10 \times 9 \times 8 \times 7 \times 6 \times 5 \times 4 \times 3 \times 2 \times 1$) different orderings and $12!$ different estimations (i.e., the sequential sum of squares) in multiple linear regression. The relative importance of each individual regressor is the mean of the $12!$ estimations. Thus, LMG is a good option for a study with a small sample size.

Results

Question 1. Do students demonstrate different temporal structures of emotions across the phases of SRL (i.e., *forethought*, *performance*, and *self-reflection*)?

As displayed in Table 3, the ratio of the repeating sequences of the same emotion (i.e., % *LAM*) significantly differed across the three phases of SRL, $\chi^2(2)=10.71$, $p=.005$. Post hoc analysis with Wilcoxon signed-rank tests was conducted with a Bonferroni correction applied, resulting in a significance level set at $p<.017$. Median (IQR) of % *LAM* at the *forethought*, *performance*, and *self-reflection* phases were 11.24 (0 to 31.25), 15.51 (4.96 to 47.47), and 23.07 (9.66 to 50.75), respectively. There were no significant differences in the % *LAM* values between the *forethought* and *self-reflection* phases ($Z = -2.12$, $p=.034$) or between the *performance* and *self-reflection* phases ($Z=-0.78$, $p=.43$). However, there was a statistically significant difference in the % *LAM* value between the *forethought* phase and *performance* phase ($Z = -3.10$, $p=.002<.016$), indicating students experienced more repeating sequences of the same emotions at the *performance* phase than *forethought* phase. However, there were no significant differences in the proportion of repeating emotions (i.e., % *REC*) across the phases of SRL ($\chi^2(2)=1.02$, $p=.60$).

Question 2. Do students have different arousal levels of emotions across the SRL phases?

The results in Table 4 showed that SCR frequency differed significantly across the three SRL phases, $\chi^2(2)=14.71$, $p=.000$. Post hoc analysis with Wilcoxon signed-rank tests was conducted with a Bonferroni correction applied, resulting in a significance level set at $p<.017$. The Median (IQR) of SCR frequency were 3.18 (1.1 to 14.47), 0.84 (0.31 to 3.61), and 2.4 (0.76 to 5.30), respectively. SCR frequency in the *forethought* phase was significantly higher than that of the *performance* phase, indicating that students had more frequent emotional arousal in the *forethought* phase. Furthermore, we found the amplitude of SCR significantly differed across the phases of SRL, $\chi^2(2)=9.74$, $p=.000$. More specifically, the amplitude of SCR in the *performance* phase was significantly lower than that of the *forethought* phase ($Z = -3.25$, $p=.001<.017$) and that of the *self-reflection* phase

Table 3 Friedman and Wilcoxon tests results of recurrence measures

Measures	χ^2	p	Phase	25th	Median	75th	Post-hoc	Z
%REC	1.02	0.60	Forethought	33.32	39.78	47.16	For. > Per.	0.26
			Performance	32.04	35.00	44.83	Per. < Ref.	-1.07
			Self-reflection	32.49	38.81	44.46	Ref. < For.	-0.51
%LAM	10.71	0.005	Forethought	0.00	11.24	31.25	For. < Per.	-3.10*
			Performance	4.96	15.51	47.47	Per. < Ref.	-0.78
			Self-reflection	9.66	23.07	50.75	Ref. > For.	2.12

Note * $p<.016$. For. = Forethought, Per. = Performance, Ref. = Self-reflection

Table 4 Friedman and Wilcoxon tests results of EDA measures

Measures	χ^2	p	Phase	25th	Median	75th	Post-hoc	Z
SCR frequency	14.71	0.000	Forethought	1.1	3.18	14.47	For. > Per.	-3.72*
			Performance	0.31	0.84	3.61	Per. < Ref.	1.25
			Self-reflection	0.76	2.40	5.30	Ref. > For.	-1.50
SCR Duration	6.32	0.042	Forethought	0.74	3.20	8.42	-	-
			Performance	1.69	5.69	21.45	-	-
			Self-reflection	0.49	4.43	6.27	-	-
SCR Amplitude	9.74	0.008	Forethought	0.77	1.96	2.49	For. > Per.	-3.25*
			Performance	0.21	0.39	0.86	Per. < Ref.	2.67*
			Self-reflection	0.35	2.58	20.81	Ref. > For.	-1.67

Note * $p < .016$. For. = Forethought, Per. = Performance, Ref. = Self-reflection

($Z=2.67$, $p=.008 < .017$). Regarding SCR duration, there was no significant difference among the three phases, $\chi^2(2)=6.32$, $p=.042 > .016$.

Question 3. What is the relative importance of the temporal structure of emotions and the arousal levels of emotions in specific SRL phases in predicting students' performance?

As aforementioned in the data analysis section, we applied Averaging over Orderings to address this question. We found that the six variables about the temporal structures of emotions (% REC – *forethought*, % REC – *performance*, % REC – *reflection*, % LAM – *forethought*, % LAM – *performance*, % LAM – *reflection*) could explain 36.82% of the variance in performance. The six variables of emotion arousal (i.e., SCR number – *forethought*, SCR number – *performance*, SCR number – *reflection*, SCR amplitude – *forethought*, SCR amplitude – *performance*, SCR amplitude – *reflection*) could explain 22.07% of the variance in performance.

When assessing the relative importance of the twelve variables to students' diagnostic performance, we found that they together account for 81.23% of the variance in performance. It is noteworthy that we aimed to assess the relative importance of each variable, instead of establishing a multiple linear regression model with those predictor variables. Therefore, we did not examine overfitting in LMG models. Meanwhile, one should be cautious when interpreting this result, i.e., the high accuracy of prediction of variance in performance. LMG is essentially not a prediction model and thus the results cannot be viewed as standardized regression coefficients. Nevertheless, the proportion of variance explained by the twelve variables did underscore the joint role of the temporal structures of emotions and the arousal levels of emotions in predicting students' performance, when compared to that explained by either alone. More specifically, the proportion of the repeating sequence of same emotions in the *forethought* phase ($R^2=0.11$) and the amplitude of SCR in the *self-reflection* phase ($R^2=0.14$) positively predicted performance, whereas the proportion of the repeating sequence of same emotions in the *self-reflection* phase ($R^2=0.25$) negatively predicted performance. These results indicate that students who had more stable emotions in the *forethought* phase, but less stable emotions and a higher level of emotional arousal in the *self-reflection* phase had better performance.

Discussion

More stable emotions in the *performance* phase

The findings of this study indicated that students experienced more stable emotions in the *performance* phase than in the *forethought* phase. A possible explanation is that the *forethought* phase, according to Schutz and Davis (2000), is “a period of mixed emotions” (p. 249). Students experienced both a variety of positive and negative emotions in the *forethought* phase. These emotions could be prospective or retrospective (Pekrun, 2006) since students started to develop an individualized understanding of the learning or problem-solving task. In contrast, students’ emotional states did not change rapidly in the *performance* phase, because they had already established an understanding of the task environment and an expectation concerning how they would approach the task (Zimmerman, 2013). For instance, students might maintain a negative emotional state if they perceived the task as being challenging at the beginning. Another explanation is that students emphasized different learning components or strategies in different SRL phases. Students concentrated on task analysis and self-motivation in the *forethought* phase, whereas they focused on the use of cognitive and metacognitive strategies to solve the task in the *performance* phase (Zimmerman, 2013). A focus on cognition and learning in the *performance* phase means that students purposely allocate efforts to control adverse behaviors, including emotion instability (Zimmerman, 2013). In the *forethought* phase, however, many factors could influence emotions in task analysis, such as individual characteristics, task difficulty, and motivation (e.g., achievement goal).

Less frequent and lower levels of emotional arousal in the *performance* phase

Moreover, this study found that students had less frequent emotional arousal in the *performance* phase than in the *forethought* phase as indicated by the significantly more SCRs in the *forethought* phase. Furthermore, students experienced a lower level of emotional arousal in the *performance* phase than in the *forethought* and *self-reflection* phases as indicated by significantly smaller amplitude of SCRs in the *performance* phase. In other words, students not only experienced relatively less frequent emotional arousal in the *performance* phase, but they were also triggered at a lower intensity. These results partially contradict the assumptions of Ainley et al. (2005), who claimed that cognitive interruption and conflicts in the *performance* phase would increase students’ physiological activities. It is possible that students control the intensity of emotions in the *performance* phase to focus on cognitive processing (Efklides et al., 2018). The less frequent and lower intensity of emotional arousal in the *performance* phase resonates with the finding that students showed more stable emotions in this phase as discussed before. During the *self-reflection* phase, however, emotions are typically more salient with higher arousal levels (Usher & Schunk, 2018). It is possible that the positive or negative feedback students received in this phase increases their levels of emotional arousal. As an illustration, students had stronger emotional responses if the feedback contradicted their beliefs (Ryan & Henderson, 2018).

Emotion stability and arousal predict student performance

In addition, we found that students who had better performance were those who demonstrated more stable emotions in the *forethought* phase, less stable emotions in the *self-reflection* phase, and a higher level of emotional arousal in the *self-reflection* phase. In the

forethought phase of self-regulated learning (SRL), we contended that the more stable students' emotional states, the higher possibility that they had confidence in solving the task. In fact, prior research has shown that students' self-efficacy affected the stability of their emotional responses, which determined students' task performance to an extent (Bandura & Freeman, 1999; Usher & Schunk, 2018). In the self-reflection phase, students reflect on the outcomes of their efforts and make attribution for what factors yield the current performance (Usher & Schunk, 2018). It is likely that students who care more about learning outcomes experience more emotional changes and a higher level of emotional arousal. For example, students were sensitive to each piece of internal and external feedback in the *self-reflection* phase so that more emotions were induced. Meanwhile, students could be either exceptionally happy when getting a correct answer or remarkably sad when making a mistake. These findings highlight the importance of analyzing the emotional changes and the level of emotional arousal in different SRL phases for providing efficient scaffoldings.

Implications

This study provides important insights for researchers to conduct fine-grained analysis of emotions with advanced emotion measures. Facial expressions and EDA data reflect different aspects of emotion attributes. While Harley et al. (2015) reported a low agreement between facial expressions and EDA, Cartaud et al. (2020) found a high correlation between them. Our findings align with Cartaud et al.'s (2020) study: specific EDA parameters (i.e., skin conduct response amplitude, skin conductance response frequency) were in line with the results of facial expressions. This indicates that the integration of these two data channels could provide a more nuanced yet coherent understanding of students' emotional responses in clinical problem-solving. Future researchers may consider concurrently employing these two methods when investigating emotion dynamics in various problem-solving contexts. Moreover, this study has methodological contributions to the analysis of fine-grained and time-series emotion data, as we took the initiative to study the temporal structures of emotions with recurrence quantification analysis (RQA) in the context of clinical problem-solving. This methodology innovation demonstrates that RQA can extend beyond the scope of group emotion synchrony, as studied by Main et al. (2016), and the daily change of emotions, as explored by Jenkins et al. (2020). Recurrence Quantitative Analysis (RQA) can be applied more extensively to time-series data to unveil the dynamics of emotions. Practically, the findings of this study inform the development of intelligent tutoring systems by suggesting the integration of technology that prompts students to use emotion regulation strategies. Wrzesien et al. (2015) offer relevant approaches for triggering emotion regulation in teenagers, which could be adapted for this purpose. As an example, forthcoming intelligent tutoring systems could encourage students to maintain a stable emotional state during task analysis, given that the proportion of the repeating sequence of the same emotions in the *forethought* positively predicted performance. Furthermore, the study offers insight into the visualization of emotions when designing intelligent tutoring systems. Facial expression and EDA data could be used to visualize students' emotion dynamics in real time. It would be interesting to examine the visualization of emotion dynamics and how it supports or hinders students' performance. It is also noteworthy that the temporal structures of emotions and the arousal levels of emotions in SRL provide important information of students'

potential performance, which significantly and meaningfully informs the research area of learning analytics.

Conclusion, limitation, and future direction

In sum, this study examined the emotion dynamics, particularly, the temporal structures of emotion and the arousal levels of emotion, in the three phases of SRL (i.e., forethought, performance, and self-reflection) in 47 students as they solved a clinical problem in an intelligent tutoring system. We used advanced emotion measures (i.e., facial expression and EDA) to capture emotions in real time. Moreover, we analyzed emotion dynamics with RQA. Findings from this study contribute to our understanding of how the features of students' emotional responses related to SRL processes and their performance.

There are several limitations to this study. First, automatic emotion recognition is not as accurate as self-report since the identification of emotions was isolated from the learning context. Although we included both facial expressions and EDA for investigation, we examined these signals separately and did not consider the convergence of these two types of data. The convergence is an important direction to address the challenges of multi-modal data. Future research should focus on examining the convergence of facial expressions, EDA, and self-reports when specific SRL events occur. Moreover, the facial expression and EDA data were collected within a single clinical problem-solving context, limiting the generalizability of findings to this specific context. While this study performed an in-depth exploration of emotion dynamics in the context of clinical problem-solving, it is essential to acknowledge that emotions can vary across different tasks. Therefore, future research should aim to replicate these findings in diverse settings. Additionally, the study is limited by its relatively small sample size, which includes participants from diverse cultural backgrounds. It will be necessary to extend the findings of this study with a larger sample to identify the patterns of emotion dynamics. Future research will also consider incorporating cultural background as a variable to explore its impact on emotion dynamics. It is also worth mentioning that EDA signals lack specificity in their meanings. For instance, SCRs are not necessarily and exclusively related to emotional arousal. Finally, we used two types of devices to collect EDA signals, although the same technique was used to process the raw EDA data. Future work will attend to these limitations to test the generalizability of our findings regarding emotion dynamics in SRL.

Author contributions

J.Z.: Conceptualization, Methodology, Writing – Original Draft. T.W.: Data Cleaning. S.L.: Data Cleaning, Formal Analysis, Reviewing, and Editing. S.P. L.: Investigation, Software, and Funding Acquisition.

Data availability

The data that support the findings of this study are available on request from the corresponding author. The data are not publicly available due to privacy or ethical restrictions.

Declarations

Competing interests

The authors declare that they have no competing interests.

Received: 27 September 2023 / Accepted: 3 April 2024

Published online: 09 May 2024

References

- Ahmed, W., van der Werf, G., Kuyper, H., & Minnaert, A. (2013). Emotions, self-regulated learning, and achievement in mathematics: A growth curve analysis. *Journal of Educational Psychology, 105*(1), 150–161. <https://doi.org/10.1037/a0030160>.
- Ainley, M., Corrigan, M., & Richardson, N. (2005). Students, tasks and emotions: Identifying the contribution of emotions to students' reading of popular culture and popular science texts. *Learning and Instruction, 15*(5), 433–447. <https://doi.org/10.1016/j.learninstruc.2005.07.011>.
- Bandura, A., & Freeman, W. H. (1999). Self-efficacy: The exercise of control. *Journal of Cognitive Psychotherapy, 13*(2). <https://doi.org/10.1891/0889-8391.13.2.158>.
- Benedek, M., & Kaernbach, C. (2010). A continuous measure of phasic electrodermal activity. *Journal of Neuroscience Methods, 190*(1), 80–91. <https://doi.org/10.1016/j.jneumeth.2010.04.028>.
- Benedek, M., & Kaernbach, C. (2010b). Decomposition of skin conductance data by means of nonnegative deconvolution. *Psychophysiology, 47*(4), 647–658. <https://doi.org/10.1111/j.1469-8986.2009.00972.x>.
- Benner, P., Hughes, R. G., & Sutphen, M. (2008). Clinical reasoning, decisionmaking, and action: Thinking critically and clinically. In Hughes RG. (eds) *Patient Safety and Quality: An Evidence-based Handbook for Nurses*. <https://www.ncbi.nlm.nih.gov/books/NBK2643/>
- Bi, J. (2012). A review of statistical methods for determination of relative importance of correlated predictors and identification of drivers of consumer liking. *Journal of Sensory Studies, 27*(2), 87–101. <https://doi.org/10.1111/j.1745-459X.2012.00370.x>.
- Braithwaite, J. J., Watson, D. G., Jones, R., & Rowe, M. (2013). A guide for analyzing electrodermal activity (EDA) & skin conductance responses (SCRs) for psychological experiments. *Psychophysiology, 49*(1), 1017–1034.
- Buono, S., Zdravkovic, A., Lasic, M., & Woodruff, E. (2020, November). The effect of emotions on self-regulated-learning (SRL) and story comprehension in emerging readers. In *frontiers in Education*. *Frontiers Media SA, 5*, 588043. <https://doi.org/10.3389/educ.2020.588043>.
- Cartaud, A., Ott, L., Iachini, T., Honoré, J., & Coello, Y. (2020). The influence of facial expression at perceptual threshold on electrodermal activity and social comfort distance. *Psychophysiology, 57*(9), e13600. <https://doi.org/10.1111/psyp.13600>.
- Carver, C. S., & Scheier, M. F. (1998). *On the self-regulation of behavior*. Cambridge University Press.
- Chentsova-Dutton, Y. E., & Tsai, J. L. (2010). Self-focused attention and emotional reactivity: The role of culture. *Journal of Personality and Social Psychology, 98*(3), 507–519. <https://doi.org/10.1037/a0018534>.
- Coco, M. I., & Dale, R. (2014). Cross-recurrence quantification analysis of categorical and continuous time series: An R package. *Frontiers in Psychology, 5*, 1–14. <https://doi.org/10.3389/fpsyg.2014.00510>.
- Dale, R., Warlaumont, A. S., & Richardson, D. C. (2011). Nominal cross recurrence as a generalized lag sequential analysis for behavioral streams. *International Journal of Bifurcation and Chaos in Applied Sciences and Engineering, 21*, 1153–1161. <https://doi.org/10.1142/S0218127411028970>.
- Dan-Glauser, E. S., & Gross, J. J. (2013). Emotion regulation and emotion coherence: Evidence for strategy-specific effects. *Emotion, 13*, 832. <https://doi.org/10.1037/a0032672>.
- Dharmathilaka, A. V. H., Jayathilaka, H. A. D. T. T., Madushanki, K. H. H. C., Jayasinghe, U., Ragel, R. G., & Bandara, D. L. (2023). Virtual patient simulator for skill training in dentistry and investigation of its effectiveness. In *2023 Moratuwa Engineering Research Conference (MERCCon)* (pp. 195–200). IEEE. <https://doi.org/10.1109/MERCCon60487.2023.10355407>
- Dindar, M., Järvelä, S., & Haataja, E. (2020). What does physiological synchrony reveal about metacognitive experiences and group performance? *British Journal of Educational Technology, 51*(5), 1577–1596. <https://doi.org/10.1111/bjet.12981>.
- Efklides, A. (2011). Interactions of metacognition with motivation and affect in self-regulated learning: The MASRL Model. *Educational Psychologist, 46*(1), 6–25. <https://doi.org/10.1080/00461520.2011.538645>.
- Efklides, A., Schwartz, B. L., & Brown, V. (2018). Motivation and affect in self-regulated learning: Does metacognition play a role? In D. H. Schunk, & J. A. Greene (Eds.), *Handbook of self-regulation of learning and performance* (pp. 64–82). Routledge/Taylor & Francis Group.
- Ekman, P., & Friesen, W. V. (1978). *Facial Action Coding System (FACS)* [Database record]. APA PsycTests. <https://doi.org/10.1037/t27734-000>.
- Elstein, A. S., Shulman, L. S., & Sprafka, S. A. (1978). *Medical problem solving: An analysis of clinical reasoning*. Harvard University Press.
- Eva, K. W. (2005). What every teacher needs to know about clinical reasoning. *Medical Education, 39*(1), 98–106. <https://doi.org/10.1111/j.1365-2929.2004.01972.x>.
- Goetz, T., Becker, E. S., Bieg, M., Keller, M. M., Frenzel, A. C., & Hall, N. C. (2015). The glass half empty: How emotional exhaustion affects the state-trait discrepancy in self-reports of teaching emotions. *Plos One, 10*(9), 1–14. <https://doi.org/10.1371/journal.pone.0137441>.
- Greco, A., Valenza, G., & Scilingo, E. P. (2016). Evaluation of CDA and CvxEDA Models. In *Advances in Electrodermal Activity Processing with Applications for Mental Health* (pp. 35–43). Springer International Publishing. https://doi.org/10.1007/978-3-319-46705-4_3.
- Grömping, U. (2006). Relative importance for linear regression in R: The package relaimpo. *Journal of Statistical Software, 17*(1), 1–27. <https://doi.org/10.18637/jaa.v017.i01>.
- Haataja, E., Malmberg, J., Järvelä, S., & Kirschner, P. (2016, November). *How is electrodermal activity related to self-regulated learning? - A case study in collaborative learning context* [Poster presentation]. Online Measures of Learning Processes, Finland. <https://doi.org/10.13140/RG.2.2.30264.96006>.
- 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*, 615–625. <https://doi.org/10.1016/j.chb.2015.02.013>.
- Horvers, A., Tombing, N., Bosse, T., Lazonder, A. W., & Molenaar, I. (2021). Detecting emotions through electrodermal activity in learning contexts: A systematic review. *Sensors (Basel, Switzerland), 21*(23), 7869. <https://doi.org/10.3390/s21237869>.
- Houben, M., Van Den Noortgate, W., & Kuppens, P. (2015). The relation between short-term emotion dynamics and psychological well-being: A meta-analysis. *Psychological Bulletin, 141*(4), 901. <https://doi.org/10.1037/a0038822>
- iMotions. (2018). *Attention Tool*. iMotions Inc.
- Isen, A. M., Rosenzweig, A. S., & Young, M. J. (1991). The influence of positive affect on clinical problem solving. *Medical Decision Making, 11*(3), 221–227. <https://doi.org/10.1177/0272989X9101100313>.

- Jenkins, B. N., Hunter, J. F., Richardson, M. J., Conner, T. S., & Pressman, S. D. (2020). Affect variability and predictability: Using recurrence quantification analysis to better understand how the dynamics of affect relate to health. *Emotion, 20*(3), 391–402. <https://doi.org/10.1037/emo0000556>.
- Kapoor, A., Burleson, W., & Picard, R. W. (2007). Automatic prediction of frustration. *International Journal of Human-Computer Studies, 65*, 724–736. <https://doi.org/10.1016/j.ijhcs.2007.02.003>.
- Kazi, H., Haddawy, P., & Suebnukarn, S. (2013). Clinical reasoning gains in medical PBL: An UMLS based tutoring system. *Journal of Intelligent Information Systems, 41*, 269–284. <https://doi.org/10.1007/s10844-013-0244-9>
- Kuiper, R. (2013). Integration of innovative clinical reasoning pedagogies into a baccalaureate nursing curriculum. *Creative Nursing, 19*(3), 128–139. <https://doi.org/10.1891/1078-4535.19.3.12>.
- Lajoie, S. P., Zheng, J., Li, S., Jarrell, A., & Gube, M. (2021). Examining the interplay of affect and self regulation in the context of clinical reasoning. *Learning and Instruction, 72*, 101219. <https://doi.org/10.1016/j.learninstruc.2019.101219>
- Lajoie, S. P., Li, S., & Zheng, J. (2021). The functional roles of metacognitive judgement and emotion in predicting clinical reasoning performance with a computer simulated environment. *Interactive Learning Environments, 1–12*. <https://doi.org/10.1080/10494820.2021.1931347>
- Lewinski, P., Uyl, D., T. M., & Butler, C. (2014). Automated facial coding: Validation of basic emotions and FACS AUs in FaceReader. *Journal of Neuroscience Psychology and Economics, 7*(4), 227. <https://doi.org/10.1037/npe0000028>.
- Li, S., Zheng, J., Lajoie, S. P., & Wiseman, J. (2021). Examining the relationship between emotion variability, self-regulated learning, and task performance in an intelligent tutoring system. *Educational Technology Research and Development, 69*, 673–692. <https://doi.org/10.1007/s11423-021-09980-9>
- Lichtwarck-Aschoff, A., Hasselman, F., Cox, R., Pepler, D., & Granic, I. (2012). A characteristic destabilization profile in parent-child interactions associated with treatment efficacy for aggressive children. *Nonlinear Dynamics Psychology and Life Sciences, 16*, 353–379.
- Lideman, R., Merenda, P., & Gold, R. (1980). *Introduction to bivariate and multivariate analysis*. Scott Foresman.
- Loijens, L., Krips, O., van Kuilenburg, H., den Uyl, M., & Ivan, P. (2015). *FaceReader 6.1. Reference manual*. Noldus Information Technology.
- Main, A., Paxton, A., & Dale, R. (2016). An exploratory analysis of emotion dynamics between mothers and adolescents during conflict discussions. *Emotion, 16*(6), 913. <https://doi.org/10.1037/emo0000180>.
- Makowski, D. (2020). Neurokit Documentation. NeuroKit. https://neurokit.readthedocs.io/_/downloads/en/latest/pdf/.
- Makowski, D., Pham, T., Lau, Z. J., Brammer, J. C., Lespinasse, F., Pham, H., & Chen, S. H. (2021). NeuroKit2: A Python toolbox for neurophysiological signal processing. *Behavior Research Methods, 53*(4), 1689–1696. <https://doi.org/10.3758/s13428-020-01516-y>.
- Marwan, N., Wessel, N., Meyerfeldt, U., Schirdewan, A., & Kurths, J. (2002). Recurrence-plot-based measures of complexity and their application to heart-rate-variability data. *Physical Review, 66*(2), 26702. <https://doi.org/10.1103/PhysRevE.66.026702>.
- Meinecke, A. L., Handke, L., Mueller-Frommeyer, L. C., & Kauffeld, S. (2020). Capturing non-linear temporally embedded processes in organizations using recurrence quantification analysis. *European Journal of Work and Organizational Psychology, 29*(4), 483–500. <https://doi.org/10.1080/1359432X.2019.1658624>.
- Menart-Houtermans, B., Rütter, R., Nowotny, B., Rosenbauer, J., Koliaki, C., Kahl, S., Simon, M. C., Szendroedi, J., Schloot, N. C., & Roden, M. (2014). Leukocyte profiles differ between type 1 and type 2 diabetes and are associated with metabolic phenotypes: Results from the German diabetes study (GDS). *Diabetes Care, 37*(8), 2326–2333. <https://doi.org/10.2337/dc14-0316>.
- Muis, K. R., Chevrier, M., & Singh, C. A. (2018). The role of epistemic emotions in personal epistemology and self-regulated learning. *Educational Psychologist, 53*(3), 165–184. <https://doi.org/10.1080/00461520.2017.1421465>.
- Niculescu, A. C., Tempelaar, D. T., Dailey-Hebert, A., Segers, M., & Gijsselaers, W. H. (2016). Extending the change–change model of achievement emotions: The inclusion of negative learning emotions. *Learning and Individual Differences, 47*, 289–297. <https://doi.org/10.1016/j.lindif.2015.12.015>.
- Pekrun, R. (2006). The control-value theory of achievement emotions: Assumptions, corollaries, and implications for educational research and practice. *Educational Psychology Review, 18*(4), 315–341. <https://doi.org/10.1007/s10648-006-9029-9>.
- Peterson, E. R., Brown, G. T. L., & Jun, M. C. (2015). Achievement emotions in higher education: A diary study exploring emotions across an assessment event. *Contemporary Educational Psychology, 42*, 82–96. <https://doi.org/10.1016/j.cedpsych.2015.05.002>.
- Raubaut, P. (2018). *SPYDER* [The Scientific Python Development Environment]. MIT Licence. <https://www.spyder-ide.org/>.
- Reitsema, A. M., Jeronimus, B. F., van Dijk, M., & de Jonge, P. (2022). Emotion dynamics in children and adolescents: A meta-analytic and descriptive review. *Emotion, 22*(2), 374. <https://doi.org/10.1037/emo0000970>
- Richardson, M. J., Dale, R., & Marsh, K. L. (2014). Complex dynamical systems in social and personality psychology: Theory, modeling, and analysis. In H. T. Reis, & C. M. Judd (Eds.), *Handbook of Research methods in Social and Personality psychology* (pp. 253–282). Cambridge University Press.
- Ryan, T., & Henderson, M. (2018). Feeling feedback: Students' emotional responses to educator feedback. *Assessment & Evaluation in Higher Education, 43*(6), 880–892. <https://doi.org/10.1080/02602938.2017.1416456>.
- Schutz, P. A., & Davis, H. A. (2000). Emotions and self-regulation during test taking. *Educational Psychologist, 35*(4), 243–256. https://doi.org/10.1207/S15326985EP3504_03.
- Shukla, J., Barreda-Angeles, M., Oliver, J., Nandi, G. C., & Puig, D. (2019). Feature extraction and selection for emotion recognition from electrodermal activity. *IEEE Transactions on Affective Computing, 20*(19), 201673. <https://doi.org/10.1109/TAFAC.2019.201673>.
- Suebnukarn, S., & Haddawy, P. (2006). Modeling individual and collaborative problem-solving in medical problem-based learning. *User Modeling and User-Adapted Interaction, 16*, 211–248. <https://doi.org/10.1007/s11257-006-9011-8>.
- Taub, M., Sawyer, R., Lester, J., & Azevedo, R. (2020). The impact of contextualized emotions on self-regulated learning and scientific reasoning during learning with a game-based learning environment. *International Journal of Artificial Intelligence in Education, 30*(1), 97–120. <https://doi.org/10.1007/s40593-019-00191-1>.
- Tomko, R. L. (2015). *Ambulatory assessment of physiological arousal, emotion, and alcohol use* (Doctoral dissertation, University of Missouri–Columbia). <https://doi.org/10.32469/10355/50193>.
- Usher, E. L., & Schunk, D. H. (2018). Social cognitive theoretical perspective of self-regulation. In D. H. Schunk, & J. A. Greene (Eds.), *Handbook of self-regulation of learning and performance* (pp. 19–35). Routledge/Taylor & Francis Group.
- van Dooren, M., de Vries, J. J. G., & Janssen, J. H. (2012). Emotional sweating across the body: Comparing 16 different skin conductance measurement locations. *Physiology and Behavior, 106*, 298–304. <https://doi.org/10.1016/j.physbeh.2012.01.020>.

- Wallot, S. (2017). Recurrence quantification analysis of processes and products of discourse: A tutorial in R. *Discourse Processes*, 54(5–6), 382–405. <https://doi.org/10.1080/0163853X.2017.1297921>.
- Wang, C. Y., Chen, S., & Huang, M. Y. (2023). Exploring medical students' metacognitive and regulatory dimensions of diagnostic problem solving. *Medical Education Online*, 28(1), 2210804. <https://doi.org/10.1080/10872981.2023.2210804>.
- Wrzesien, M., Rodríguez, A., Rey, B., Alcañiz, M., Baños, R. M., & Vara, M. D. (2015). How the physical similarity of avatars can influence the learning of emotion regulation strategies in teenagers. *Computers in Human Behavior*, 43, 101–111. <https://doi.org/10.1016/j.chb.2014.09.024>.
- Zheng, J., Lajoie, S. P., Li, S., & Wu, H. (2023). Temporal change of emotions: Identifying academic emotion trajectories and profiles in problem-solving. *Metacognition and Learning*, 18(2), 315–345. <https://doi.org/10.1007/s11409-022-09330-x>
- Zheng, J., Li, S., Lajoie, S.P. (2023). A review of measurements and techniques to study emotion dynamics in learning. In: Kovanovic, V., Azevedo, R., Gibson, D.C., Ifenthaler, D. (eds) *Unobtrusive Observations of Learning in Digital Environments. Advances in Analytics for Learning and Teaching*. Springer, Cham. https://doi.org/10.1007/978-3-031-30992-2_2
- Zimmerman, B. J. (1990). Self-regulated learning and academic achievement: An overview. *Educational Psychologist*, 25(1), 3–17. https://doi.org/10.1207/s15326985ep2501_2
- Zimmerman, B. J. (2000). Attaining self-regulation: A social cognitive perspective. In Boekaerts M, Pintrich P. R, & M. Zeidner (Eds.), *Handbook of Self-Regulation* (pp. 13–39). Academic Press. <https://doi.org/10.1016/B978-012109890-2/50031-7>.
- Zimmerman, B. J. (2013). From cognitive modeling to self-regulation: A social cognitive career path. *Educational Psychologist*, 48(3), 135–147. <https://doi.org/10.1080/00461520.2013.794676>.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.