Affective Dynamics and Cognition During Game-Based Learning

Elizabeth B. Cloude, Daryn A. Dever, Debbie L. Hahs-Vaughn, Andrew J. Emerson, Roger Azevedo, and James Lester

Abstract—Inability to regulate affective states can impact one's capacity to engage in higher-order thinking like scientific reasoning with game-based learning environments. Many efforts have been made to build affect-aware systems to mitigate the potentially detrimental effects of negative affect. Yet, gaps in research exist since accurately capturing and modeling affect as a state that changes dynamically over time is methodologically and analytically challenging. In this paper, we calculated multilevel mixed effects growth models to assess whether seventy-eight participants' (n = 78) time engaging in scientific reasoning (via logfiles and eye gaze) were related to time facially expressing confused, frustrated, and neutral states (via facial recognition software) during game-based learning with Crystal Island. The fitted model estimated significant positive relations between the time learners facially expressed confusion, frustration, and neutral states and time engaging in scientific-reasoning actions. The time individual learners facially expressed frustrated, confused, and neutral states explained a significant amount of variation in time engaging in scientific reasoning. Our finding emphasize that individual differences and agency may play a important role on relations between affective states, their dynamics, and higher-order cognition during game-based learning. Designing affect-aware game-based learning environments that track the dynamics within individual learners' affective states may best support cognition.

Index Terms—Game-based learning environments, multimodal data, cognitive trends, affective dynamics

1 Introduction

Affect is a feel, emotion, or mood represented by cognitive structures in the mind that provide information about the world, compelling us to take action and make decisions [1], [2]. Across scientific communities, significant efforts are being made to build intelligent systems capable of automatically detecting and mitigating harmful affective states [3], [4], [5]. For example, imagine affect-aware systems capable of supporting learners in regulating confusion while scientifically reasoning during game-based learning. However, progress is slow due to a heavy reliance on data-driven techniques that often ignore the cognitive basis affective states [6]. In this study, we addressed this challenge by adopting a perspective in cognitive-affective theory of learning that explains relations between affect and cognition with emerging technologies such as game-based learning environments [7]. We converged multiple channels of time series data including facial expressions of emotions, eye gaze, and learner-system interactions to examine relationships between affective dynamics and trends in cognition (e.g., scientific reasoning) during game-based learning. The findings of this work suggest directions for designing affect-aware systems to not only track and monitor how long affective states persist, but also emphasize how the system can best intervene when affect suggests detrimental relations with cognitive abilities (e.g., inability to reason due to negative affect) based on individual characteristics (e.g., lack of prior knowledge) during game-based learning.

- *Elizabeth B. Cloude is with the Penn Center of Learning Analytics, University of Pennsylvania, Philadelphia, PA 19104 USA. E-mail: ecloude@upenn.edu.*
- *Daryn A. Dever and Roger Azevedo are with the School of Modeling, Simulation, and Training, University of Central Florida, Orlando, FL 32816 USA. E-mail: ddever@knights.ucf.edu, roger.azevdeo@ucf.edu.*
- *Debbie L. Hahs-Vaughn is with the College of Community Innovation and Education, University of Central Florida, Orlando, FL 32816 USA. E-mail: debbie@ucf.edu.*
- *Andrew J. Emerson is with Multimodal AI Team, Educational Testing Service, Princeton, NJ 08541 USA. E-mail: aemerson@ets.org.*
- *James Lester is with the Department of Computer Science, North Carolina State University, Raleigh, NC 27606 USA. E-mail: lester@ncsu.edu.*

Manuscript received 20 January 2022; revised 16 September 2022; accepted 19 September 2022. Date of publication 29 September 2022; date of current version 15 November 2022.
This work was supported in part by the Social Sciences and Humanities Research Council of Canada under Grant SSHRC 895–2011–1006.
This work involved human subjects or animals in its research. Approval of all ethical and experimental procedures and protocols was granted by North Carolina State University Institutional Review Board for the Use of Human Subjects in Research.
(Corresponding author: Elizabeth B. Cloude.)
Recommended for acceptance by M. T. Sezgin.
Digital Object Identifier no. 10.1109/TAFFC.2022.3210755

2 Game-Based Learning Environments

Game-based learning environments are designed with game elements to enhance cognition by fostering affective experiences that lead to engagement, and thus are key research tools for studying relations between affective dynamics and cognition. A systematic review by [8] found that out of 149 studies, game-based learning environments not only influenced perceived control and value in the learning experience, but that learners also reported more emotional engagement during learning compared to traditional education settings [7]. Several other meta-analyses support the same conclusion: game-based learning environments are useful tools to enhance cognitive skills due to their affect-inducing features compared to other education settings like classrooms [9], [10].

A game-based learning environment called Crystal Island, for example, fosters affect and engagement by manipulating level of agency (i.e., restricted versus total autonomy) afforded to learners in how they can interact with game features. Studies empirically support that the game elements built into Crystal Island foster cognitive processing that allow learners to reason scientifically [11]. In other words, Crystal Island presents realistic scenarios with clearly defined goals linked to structured learning activities. Learners with restricted agency must follow a pre-set sequence of learning activities, while learners in the total agency condition are free to define their own sequence of learning activities. Studies empirically support that manipulating agency in Crystal Island fosters more cognitive processing, engaging learners in scientific reasoning and enhancing their learning of content [11].

However, these research findings present key challenges as most of the studies failed to account for the cognitive basis of affect and captured affect (e.g., engagement) using self-report data or observations. While these methods are valid, we argue that in order to measure affect, researchers also need continuous streams of information on affect to capture its stability or change over time. Self-report data and observations reveal affect at static intervals, potentially missing information on affective dynamics and its relation to cognition and learning with game-based learning environments over time [6]. We argue that utilizing these methods fails to acknowledge the changing nature of affect, ignoring critical theoretical assumptions. In this paper, we addressed these issues by referring to the integrated model of cognitive-affective learning with media (ICALM; [7]) and utilizing continuous streams of facial expressions. We argue this approach might offer more insight in to relations between affective dynamics and cognition during game-based learning.

We argue that Crystal Island is an ideal platform for studying relations between affect and cognition guided by ICALM because it was built with features that manipulated the level of agency learners had, influencing their affect and cognitive processing. For instance, one study [12] examined the role of agency during learning with Crystal Island and its impact on affect, motivation, and learning outcomes. Results showed that learners with restricted agency during game-based learning achieved the highest learning gain compared to the total and no agency conditions. They also found that learners in both the no and total agency conditions expressed more frustration and confused affective states during learning activities compared to the restricted agency condition. Although, challenges remain since the study aggregated all instances of affect, missing how affect might have changed or persisted over time and its relation to cognition across the agency conditions. We address these challenges in our study using time series data to analyze longitudinal trends in affective states and its relation to scientific reasoning across agency conditions during game-based learning with Crystal Island.

3 Integrated Model of Cognitive-Affective Learning With Media

ICALM assumes two components exist when learners cognitively process information with emerging technologies: (1) working memory and (2) long-term memory using two channels–verbal and nonverbal [13]. Working memory is responsible for processing and integrating new information into long-term memory and has a very limited storage capacity and duration (i.e., cognitive load [13]. Conversely, long-term memory permanently stores information at an unlimited capacity in memory, where this information can bypass working memory if it is like other information already stored in long-term memory using chunking mechanisms (i.e., prior knowledge). The ICALM framework also explains that there is a third channel that moderates working memory while cognitively processing information into long-term memory: affective states [14]. Affective states are an umbrella term that describe feelings, moods, attitudes, and emotions, that range from suffering to elation. Consisting of multiple dimensions that include psychological, cognitive, motivational, and physiological components , affect leads to a pleasant or unpleasant subjective experience (i.e., positive or negative; valence [15]).

According to ICALM, cognitive processes are emerging properties that result from interacting prior knowledge, abilities, beliefs, affect, and motivation. As such, affect cannot be decoupled from cognitive processes and it controls whether goal-directed behaviors like scientific reasoning are initiated and sustained during game-based learning, either consciously or unconsciously [16]. While a learner is constructing and organizing information via working memory, it may induce affect (e.g., frustration or confusion if prior knowledge contradicts new information). Further, studies utilizing ICALM suggests that positive affective states (e.g., engagement) were related to better cognitive capacity and information processing with game-based learning environments [17].

However, relations are less clear for negative affective states and cognitive processes. ICALM argues that the maximum cognitive capacity individual learners have impacts their cognitive processing rather than the amount of cognitive resources used. For instance, a task that is too easy or too difficult will induce affective states, and depending on the affective state and whether the learner can resolve it, determines whether the learner's motivation is undermined and if cognitive processing is sustained. Since the learner is working beyond their current abilities, it is possible that they will not devote enough cognitive resources to the learning activity. We discuss empirical results on affect and cognition with emerging technologies in the following section.

3.1 Studying Affective States and Cognition

Studies find inconsistent relations between negative affect and cognitive abilities with emerging technologies, where sometimes negative affect is beneficial to cognitive abilities, and other times it is harmful [15], [18], [19], [20]. We argue that these mixed findings may result from a lack of theoretical integration and common methodologies for measuring affect that involve administering self-report questionnaires measuring affect before or after learning activities [21], and more recently in-situ using facial recognition and deep learning approaches [22]. Other studies converge self-report items with other data channels such as learner-system interactions [2]. For instance, [2] administered self-report items every 14 minutes during learning activities to capture the valence, dynamics, and activation of affective states over time [23]. They aligned the self-report data with log files indicating cognitive strategy use (e.g., summarizing content) to gain insight into affect and its relation to cognitive processing. Latent growth models showed that when negative activating affective states did not change over time, it was related to cognitive strategy use. They also found that relationships between changes in negative activating affective states and the quality of cognitive strategies predicted the highest and lowest 30% learners. This finding suggested that studying affective states defined by valence, activation, and dynamics could provide deeper insight into its relation with cognition and its association to learning outcomes with emerging technologies.

However, progress is slow because affect is typically measured at static intervals. For instance, a study by [24] found that some self-reported negative affective states were associated with *more* cognitive processing [20], [25], while other negative affective states were associated with less. When learners experienced confusion, it resulted in more engagement with the content and thus more cognitive processing of information [24]. Another study by [26] found that when learners experienced prolonged bored boredom, it was associated with poor problem-solving performance with emerging technologies; yet, when learners reported frustration that changed across learning activities, it was associated with better problem-solving performance [26].

A similar study by [27] examined whether, and to what extent, changes in self-reported bored, frustrated, and confused affective states captured using self-reports over the course of learning activities impacted self-regulation and performance [27]. They found that negative affective changes impacted cognition. But, these relations were not consistent across all negative affective states. Different relationships were present with cognition based on which and when a particular affective state changed during the learning session. For example, increases in boredom were related to lower performance, while no changes in confusion were related to less learning. Another study examined affective dynamics and their relationship to changes in self-regulation and achievement using growth curve analysis [28]. Their models suggested more dynamic and complex relations between negative affective dynamics, self-regulation, and achievement. Specifically, they found variability in how individual's negative affect changed over learning activities, suggesting that individual differences may be moderating relations between affect and cognition. This finding is aligned with ICALM assumptions which suggest that the maximum cognitive capacity is defined by the learners' affect and individual characteristics. Although, it should be emphasized that these studies utilized static interval of affect using self-report items. Future research calls for continuous streams of data on affect to detect more granular changes over time.

Observational protocols are another method for studying affective states over time, particularly during classroom learning (e.g., BROMP [25]) that provide more granularity compared to self-report items. Within BROMP, for example, learners are individually tagged in a pre-determined order using a momentary time sampling method. Field observers are trained on how to detect a representative sample of behavior and affect based on body language, facial expressions, and many other indicators of affective states. A study by [29] leveraged BROMP observations to assess the transition of affective states and their relation to learning outcomes. They found that the frequency of boredom patterns was a powerful indicator of learners' prior knowledge, yet not an indicator of learning. At present, the most granular method used today is collecting affect using facial recognition software. The Facial Action Coding System (FACS [30]) is a technique the offers a comprehensive approach for identifying all visible muscle movements on the face and are being used in facial recognition software to automatically detect affect [31]. A study by [32] used FACET, a facial recognition software built off FACS, to examine the role of negative affective states on self-regulation and problem solving with Crystal Island. Specifically, they defined affect by converging facial expressions with context (e.g., facial expressions during specified events) during game-based learning. Results showed that the context in which affect was facially expressed (e.g., confusion following experimental testing versus confusion while reading scientific articles) played a role on cognitive processing [32]. However, limitations exist as the facial expressions data were aggregated and averaged over time, removing information on the stability or change in affect over time. For instance, statistical models built to handle time series data allow researchers to extract trends in affective states to study if, when, and how affect changes and its role on cognitive processing rather than as discrete constructs (e.g., static or aggregated). To develop a better understanding on relations between affective dynamics and cognition, it is essential to leverage statistical models and continuous time series data on affect and cognition guided by a contemporary theoretical perspective and empirical support [33].

Overall, we argue that to address the challenges in literature we need to move towards continuous and longitudinal data on affective states and build statistical models that can extract trends. In this paper, we focused on negative affective states, confusion and frustration, that are outlined in the Affective Dynamics Model [24] because these affective states indicate when a learner has reached an impasse in their cognitive processing of information. If a learner remains in a confused state, this model indicates that confusion will transition into a frustrated state revealing that they might be stuck and cannot cognitively process information [2], [23], [24], [26], [27], [28], [29]. We emphasize that the affective dynamics between confused and frustrated states in relation to cognitive processing will provide insight into if and when negative affect might be detrimental to cognitive processing of information.

4 Current Study

The objective of this article was to address significant gaps in studies examining relationships between confusion, frustration, and neutral affective dynamics, as well as time engaging in cognitive activities over the course of related to scientific reasoning game-based learning with Crystal Island across two experimental conditions (restricted versus total agency groups). Through leveraging facial recognition software, we captured the time learners' facially expressed frustrated and confused affective states during game-based learning. We also collected time-stamped learner-system interactions and converged these data with eye-gaze fixations as a measure of engagement with materials and generated latent variables of scientific reasoning. For example, when learners interacted with game elements that might suggest engaging in scientific reasoning (e.g., gathering information through reading a book), we collected the amount of time their eyes fixated on the material as well as the time the log files indicated their interaction with the game element. We guided our research using ICALM [7] and built multilevel growth models to account for affective dynamics [27], [28], [32]. Our research questions are outlined below:

4.1 To What Extent Does Time Scientifically Reasoning During Game-Based Learning Change Over Time

We hypothesize learners will demonstrate changes in the time spent scientifically reasoning over the course of game-based learning based on the ICALM and empirical literature [7], [19]. Specifically, we hypothesize that time scientifically reasoning changes over time

4.2 To What Extent are There Relationships Between Time Expressing Frustration, Confusion, and Neutral States and Time Scientifically Reasoning During Game-Based Learning, and do These Relationships Differ Between Experimental Conditions?

We expect there will be relationships between time expressing confused, frustrated, and neutral states and time scientifically reasoning based on previous empirical evidence and ICALM [7], [17]. Specifically, we hypothesize that relationships will exist between confused, frustrated, and neutral states as well as time scientifically reasoning, and that these relationships will differ between restricted and total agency groups during game-based learning. We do not state a directional hypothesis since little to no research has been done to investigate the role of agency on relationships between affective dynamics and trends in scientific reasoning with game-based learning.

5 Methods

5.1 Participants

A sample of 138 undergraduate students were recruited for a large study across public institutions in North America to solve a mystery with Crystal Island, a narrative-centered game-based learning environment designed to foster (1) higher-order thinking skills (e.g., scientific reasoning) and (2) microbiology knowledge [12], [34]. Participants were randomly assigned to 1 of 3 experimental conditions: 1) restricted agency, 2) full agency, and 3) no agency. In this paper, we only analyzed participants who were assigned to the restricted or full agency conditions. Participants in these conditions were included in the subsample based on several criteria: (1) complete data for facial recognition, time-stamped log files, and eye gaze data channels, and (2) non-outlying data points. Specifically, we removed a total of 60 participants from the original sample if they had 1) incomplete data, 2) outliers via Grubb's test [35], and/or 3) were assigned to the no agency condition; see rationale in Section 5.3). As such, a subsample of 78 undergraduates' multimodal data (67% female; $M_{Age} = 20.03$, SD=1.71) were analyzed in this manuscript.

Most of the sample identified as 'White/Caucasian' (69%) and reported average video game skills (32%) and playing 0-2 hours per week (71%). None of the participants reported learning with Crystal Island before the study. The Institutional Review Board approved the study prior to recruitment and data collection.

5.2 Crystal Island, a Game-Based Learning Environment

Crystal Island was designed to enhance scientific-reasoning skills while learning about microbiology topics [12]. Upon starting the game, participants arrived via boat on a virtual island where they were confronted by a researcher who informed them that a mysterious pathogen had infected the research team. Participants were instructed to adopt the role of a Center for Disease Control agent and solve the mystery by identifying (1) the pathogen, (2) its transmission source, and (3) a treatment plan. For participants to solve the problem, they used resources and tools built to foster scientific reasoning such that they could gather information, generate hypotheses, and experimentally test those hypotheses. Specifically, Crystal Island provided clues via non-player characters who explain the patient symptoms or potential transmission sources. Several books, research articles, and posters were integrated throughout the game, covering information on a range of pathogens like viruses and bacteria. Participants had opportunities to gather food items in a backpack and test them on a virtual scanner to evaluate if the food was contaminated with a pathogen. When participants gathered enough evidence to make a final and accurate diagnosis, they completed the game.

5.2.1 Tools for Solving the Problem

During game play, participants had access to a tool known as the diagnosis worksheet, where they could record all clues and hypotheses they had about the potential pathogen and transmission source. The diagnosis worksheet was divided into four different sections: (1) Patient's Symptoms to document the symptoms sick patients reported, (2) Test Results to document the results obtained from scanning various food items, (3) Possible Explanations to document the likelihood of a given pathogen as the illness impacting the research team, and (4) Final Diagnosis to propose their hypothesized solution for the suspected pathogen including the pathogen, transmission source, and treatment. When participants felt they had correctly identified the final diagnosis, they submitted their solution. If it was correct, the participant successfully solved the problem, whereas if the solution was incorrect, the participant was instructed to try for another solution. Other resources built to foster scientific reasoning included disseminating information about microbiology via books, research articles, and posters throughout various building in Crystal Island using text and diagrams. After reading a book, participants had opportunities to immediately assess their understanding of the information using a tool called a concept matrix. The concept matrices required participants to match information using a matrix, such as matching treatment solutions (e.g., vaccine) with different types of pathogens (e.g., anthrax). Participants completed the concept matrix after reading a book and were given a total of 3 attempts to fill it out correctly. Participants also had opportunities to use a scanner to test their hypotheses (i.e., test food items for contamination) and store food items they hypothesized to be contaminated with a pathogen in their backpack.

5.3 Experimental Design

Crystal Island was designed to foster scientific reasoning using three experimental conditions that manipulated agency. The control condition (i.e., total agency; n=62), granted participants full control over their actions, while the experimental condition (i.e., restricted agency, n=16) required participants to follow a sequence of actions (e.g., reading books first) designed to foster scientific reasoning. The no agency condition granted participants no control over their actions while they watched in a playthrough of the game. Since this group had no agency over how they interacted with Crystal Island for this group, we did not consider their data in our sample and analysis. The no agency condition was designed to model what optimal scientific reasoning looks like during game-based learning. Upon starting the game, all participants began the tutorial that introduced the narrative and how to interact with the environment.

5.4 Procedure

Upon entering the research laboratory and obtaining written consent, a researcher instructed the participant to sit in front of a computer and they were randomly assigned to an experimental condition. Participants were run through the experiment individually to control for confounding variables. An electrodermal bracelet was placed around their wrist and they were calibrated to an SMI EYERED 250 eye tracker using a 9-pt calibration. Next, the participant was instructed to view a grey screen with a neutral expression for approximately 10 seconds to establish a baseline for the facial recognition software. Once the participant completed demographic and pre-test items, they began the tutorial with Crystal Island. On average, it took participants in the control condition 81 minutes (SD=23 minutes) to solve the mystery, while it took 78 minutes (SD=16) for participants in the experimental condition, or up to a maximum of 90 minutes when the game ended regardless of whether they solved the problem. Afterwards, participants completed a similar 21-item, multiple-choice, post-test assessment on microbiology and self-report items gauging motivation and cognitive load. The participant was then debriefed, paid \$10/hour, and thanked for their time.

5.5 Apparatus

5.5.1 Timestamped Log Files

Timestamped log files captured time participants initiated actions using the mouse and/or keyboard for analysis. Specifically, this data channel provided event- and time-based data during game-based learning, capturing the frequency and duration of all learner-system interactions. We used this channel to align and extract data from the other sources.

5.5.2 Facial Recognition Software

To capture affective states, we measured participants' facial expressions using a video-based facial expression and affect recognition tracking system known as FACET [31], [36]. FACET has been validated by empirical testing [36], [37]. Specifically, [36] conducted a systematic comparison between 14 databases with dynamic facial expressions using FACET software to assess whether it was comparable to human recognition performance. Detection rates were above 50% for the majority of the databases, but a proportion of facial stimuli could not be recognized by the machine-learning algorithm, showing significant differences in recognition performance between the databases. This suggests that the FACET software is sensitive to potential differences in facial features or structures across humans, and that diversity may impact the accuracy of emotion recognition and detection. Another study by [38] compared Affectiva, another facial recognition software, with FACET. Accuracy differed for distinct emotions, and FACET performed better. Overall, FACET achieved acceptable accuracy for standardized pictures, but performed worse for more natural facial expressions. While we acknowledge that facial recognition software has its limitations [39], we would like to highlight their value in capturing continuous, longitudinal data on facial expressions which provides an opportunity to examine affective dynamics over time compared to other measurement tools of affective states (see Section 3.1).

The facial features correspond to a defined set of facial muscle movements (i.e., action units such as an inner brow raiser), and composite affective states (i.e., a combination of different action units that correspond to a particular emotion). Specifically, FACET utilizes a support vector machine that was trained to capture facial features, based on the Facial Action Coding System (FACS [30]), when they deviated from the participant's baseline (i.e., neutral facial expression). To increase our ability to capture affective states across learners, we instructed participants to display their neutral expressions prior to engaging with Crystal Island, which served as a base line for the facial recognition software to facilitate the detection of deviations from individual baseline expressions. The software generates an evidence score for each affective state and action unit based on the degree that facial landmarks have deviated from the baseline neutral facial expression. An evidence score is generated when the deviation occurs and represents the odds of a human coder classifying the presence of an affective state or action unit deviation from baseline on a log likelihood scale. This means that when an evidence score of 1 occurs for a confused affective state, then the 1 would indicate that there is a likelihood that 10 human coders would classifying that there is a confused affective state present.

TABLE 1
Operational Definitions for Scientific-Reasoning Actions

Variables	Data channels	Game elements
Action 1: gathering information	Timestamped log files and eye fixations	Reading books, research articles, and posters; talking to NPCs.
Action 2: hypothesis generation	Timestamped log files and eye fixations	Backpack, food items, first field in the diagnosis worksheet.
Action 3: experimental testing	Timestamped log files and eye fixations	Final diagnosis field on the worksheet, concept matrix, and scanner.

For purposes of this study, we extracted and converged facial expressions of confused, frustrated, and neutral states when the evidence score indicated at least a 1 . An evidence score of 1 or higher was used as an indicator that the affective state was present.

5.5.3 Eye Tracker

An SMI EYERED 250 eye tracker [40] was used in this study to detect pupil and fovea locations via infrared light. We captured eye movements at a sampling rate of 30 Hz using a 9-point calibration while participants learned with Crystal Island. This calibration allowed us to measure eye movements at an offset of less than 0.05mm. Specifically, calibration is the process whereby the geometric characteristics of a participant's eyes are estimated as the basis for a fully-customized and accurate gaze-point calculation. The highest calibration on the eye tracker was 9-points across the screen, allowing for the most accurate estimation of the geometric shape of the eye and location of the pupil. While seated, participants were asked to observe a moving dot on the eye-tracking monitor. This calibration process took less than 1 min to complete. Next we analyzed the captured eye movement data for quality. This process was completed by examining the quality of eye movement recordings and removing the data sets for those participants that had less than 80% gaze sample. The gaze sample refers to percentage of the times that eyes were correctly detected by the eye tracker for each participant. For example, 100% means that one or both eyes were detected by the device throughout the recording; 50% means that one eye or both eyes were found for half of the recording duration. While screen-based eye tracking experiments typically require users to look at the screen while completing a task, some people may look away or look down (e.g., at the keyboard or mouse) to think about a problem. We processed the eye-tracking data using iMotions software [31] and defined Areas of Interest (AOIs) around game elements related to scientific reasoning such as a research article or book. AOIs were operationally defined as gaze points within 1° visual angle that lasted for at minimum 250 ms [41], and we used these metrics to generate the amount of time participants fixated on game elements related to scientific reasoning (i.e., actions 1-3; see Table 1).

5.6 Coding and Scoring

5.6.1 Aligning Data Channels

Data were extracted and temporally aligned using a pipeline built in Python [42]. Processing, cleaning, analysis, and data visualizations were conducted using R software version 3.6.0 [43]. We computed multilevel growth models using 'lme4' [44]. Three data channels were aligned in this study: (1) timestamped log files, (2) eye gaze, and (3) facial expressions of emotions. To combine these data on the same temporal scale, we converted eye gaze and facial expressions data from milliseconds to seconds. Next, we examined when participants initiated actions during game-based learning and relied on the log file timestamps to dictate when we would extract both eye-gaze and facial expressions variables. When a participant initiated an action (e.g., opening a book), we referenced the time elapsed to align whether an facial expression was present at the start of the action (e.g., confusion detected at the timestamp when the book was opened), as well as whether the participant was fixating on a game element while they were engaging in an action via log file (e.g., opened a book in the log file and was also fixating on the book via eye gaze). Each time a participant initiated an action that corresponded with scientific reasoning (Table 1 for more details), we combined facial expression and eye gaze data such that it corresponded to the timestamp in the log file.

5.6.2 Scientific-Reasoning Action Variables

To define scientific-reasoning action variables, we used a technique that combines both eye fixations and timestamped log files (e.g., see [45]) that corresponded to scientific-reasoning actions outlined in Table 1. Actions were considered scientific reasoning based on the scientific reasoning as dual space framework [46]. Specifically, when the log file suggested that a participant initiated a scientific-reasoning action (e.g., gathering information by talking with a non-player character), then we referenced the timestamp and assessed whether the participant was fixating on the game element that corresponded with the action (e.g., non-player character) to ensure they were engaged in the action rather than aimlessly (or accidentally) selected game elements. Specifically, if a participant opened a book 5 minutes into the game, it was reflected in the log file and we defined the action as gathering information if the participant only opened the book and fixated on the text in the book for longer than 250ms. Further, we extracted the time participants engaged in scientific-reasoning actions only when eye-gaze and log files were consistent over time. We included eye gaze because empirical evidence supports that when participants fixate on content, it has been indicative of cognitive processing of information with emerging technologies [45], [47].

5.6.3 Facial Expressions of Emotions

Affective states were defined using evidence scores that represented deviation from the participant's baseline facial expression (see 5.4.2. for details). Specifically, we extracted when facial expressions of frustration, confusion, and neutral state were captured at the start of scientific-reasoning actions (e.g., reading a book) and were computed based on the amount of time participants facially expressed these affective states during actions outlined in Table 1. Using a combination of relative and absolute thresholding of amplitude, the evidence scores were preprocessed and smoothed within an 11-step window for evidence scores representing at least 1 or above, and these scores were standardized by learner to a unit normal distribution to account for individual differences in facial expressions. The relative thresholding was calculated by classifying an affective state based on an evidence score representing a standardized value above 1.65 (the top 5% of observations) for at least 0.5 seconds to avoid capturing micro-expressions and to increase validity. Absolute thresholding was also calculated to classify events as presence of an affective state when the raw evidence score was at or above 1 to avoid values that were negative (represented the likelihood that the affective state was not present). This technique yielded continuous events during game-based learning that represented muscle contractions above learners' baseline (i.e., neutral) affective state for at least 0.5 seconds according to FACS.

5.7 Statistical Analysis

To estimate the general trend of time scientifically reasoning during game-based learning and its relation to time facially expressing frustration, confusion, and neutral states , we used multilevel growth models [48]. Multilevel growth models can handle unbalanced design or unequally spaced measurements, hierarchical structures, and hold less stringent assumptions. To estimate the general trend of time scientifically reasoning and its relation to time facially expressing frustrated, confused, and neutral states during scientific reasoning, we calculated two-level multilevel growth models of longitudinal change [27], [28], [32].

The Level-1 model (within-individuals; 15,882 units; see Equation (2)) described each learner's change in time scientifically reasoning trajectory using growth curve parameters, where the intercept represented the learner's initial status while the slope represented the learner's growth rate. The outcome variable represented by SR indicates scientific reasoning over time in Equations (1)–(4). The Level-2 model (between-individuals; 78 units) estimated learners' individual differences in growth curve parameters (i.e., inter-individual differences in scientific-reasoning change) to model whether features (e.g., duration of frustration present during scientific reasoning) of individual change trajectories vary across individuals (i.e., random effects; see *r* in Equations (3)– (4)). We also included experimental condition and pretest scores as level two predictors to examine whether these predictors varied between individuals, potentially indicating a moderation on relationships between time facially expressing frustrated, confused, and neutral states and time scientifically reasoning during game-based learning.

First, we estimated an unconditional means model (i.e., null model with no predictors) to partition the within- and between-individual variance over time scientifically reasoning by obtaining the intra-class correlation coefficient (ICC). ICC measures the total variance explained between individual learners. Next, to estimate the general trend of time scientifically reasoning during game-based learning, we estimated an unconditional growth model (with no predictors other than game play time [48]). We used these initial models to determine whether there were systematic mean level change as well as individual variability in change related to time scientifically reasoning over game play.

For the second research question, we calculated three separate two-level multilevel growth models to examine relationships between predictors: facially expressing (1) confused, (2) frustrated, and (3) neutral states on outcome variable: trend of time scientifically reasoning during game-based learning at an alpha level of 0.05. All predictors were defined as fixed effects, and we added a random effect for each individual to assess the extent to which variation between predictor and outcome variables (i.e., trend of time scientific reasoning) was explained by individual differences. We assessed model fit using deviance metrics, likelihood ratio test [49], and pseudo R^2 [48]. We grand-mean centered continuous predictor variables representing time facially expressing frustration, confusion, and neutral states. These variables were centered because this technique moves the mean to zero to standardize the magnitude of time (e.g., X1 – Average) to compute meaningful intercepts that reflect values when predictors were at zero. The time variables for facial expressions of frustrated, confused, and neutral states served as predictors and were centered to the start of each scientific-reasoning action. We also computed time within the game, where game play time was centered immediately after the tutorial was completed in Crystal Island so that we only analyzed time that learners were interacting with the game rather than completing the tutorial. Centering thus helps us to establish meaningful zero points which, in turn, affects our regression output. We selected this centering technique to compute an intercept that represented a time of 0 (at the start of learning with Crystal Island). To test the statistical significance of random effects, we used Type III Sum of Squares because it is not sample size dependent and was built to handle unstructured time variables [50], such as game time which varied between individuals. Full maximum likelihood estimation was used since we compared models that differed in both fixed and random components throughout model building stages.

Unconditional Models:

$$Y_{\mathrm{ij}} = \lambda_{00} + \lambda_{01}{}^{*}(Time)_{\mathrm{ij}} + \sigma_{0\mathrm{j}} + \sigma_{1\mathrm{j}}{}^{*}(Time)_{\mathrm{ij}} e_{\mathrm{ij}} \tag{1}$$

where Y_{ij} describes the outcome variable (e.g., time scientific reasoning), lambda$_{00}$ and λ_{01} are, respectively, mean initial status and average growth rate. The symbols σ_{0j}, σ_{01}, and e represent, respectively, residual variance in initial status, residual variance in growth rate, and within-person residual variance.

Condition Models:

$$\begin{aligned} Y_{\mathrm{ij}} &= \lambda_{00} + \lambda_{01}{}^{*}(Time) + \lambda_{02}{}^{*}(Action) \\ &+ \lambda_{03}{}^{*}(FacialExpression) \\ &+ \lambda_{04}{}^{*}(FacialExpression)^{*}(Time) \\ &+ \lambda_{05}{}^{*}(PreTestScore) + \lambda_{06}{}^{*}(ExperimentalCondition) \\ &+ \sigma_{0\mathrm{j}} + \sigma_{1\mathrm{j}} + e_{\mathrm{ij}} \end{aligned} \tag{2}$$

TABLE 2
Descriptive Statistics of Variables

Variables	Experimental $M(SD)_1$	Control $M(SD)_0$
Action 1		
Overall	2447.75 (415.79)	1880.29 (702.61)
Frustration	1142.7 (792.5)	1837.12 (124.35)
Confusion	1420.65 (111.2)	1772.8 (123.79)
Neutral	2242.9 (121.7)	2034.1 (170.9)
Action 2		
Overall	509.60 (318.41)	433.83 (294.34)
Frustration	2387.69 (181.6)	3786.44 (344.70)
Confusion	3004.41 (268.60)	3680.04 (338.73)
Neutral	4368.25 (288.23)	3986.77 (411.67)
Action 3		
Overall	454.54 (137.07)	420.22 (199.49)
Frustration	1521.14 (112.52)	2198.85 (191.16)
Confusion	1954.44 (180)	2155.51 (197.20)
Neutral	3008.83 (192.06)	2328.81 (259.06)

Note. Action 1=Gathering information; Action 2=Generating hypotheses; Action 3=Experimental testing; data are in seconds.

TABLE 3
General Trend of Time Changes in Scientific-Reasoning Actions

	Null Estimate (SE)	Growth Estimate (SE)	Mixed Model Estimate (SE)
Fixed effects			
Initial status	31.48^*(0.90)	49.55^*(1.63)	51.06^*(2.54)
Growth rate		-0.01^*(0.00)	-0.01^*(0.00)
Random effects			
Within-individual			68.06^*
Initial status			19.93^*
Growth rate			0.004^*
Covariance			-0.71
Deviance	179732	179406	179356
ICC	0.01	0.02	0.08

*Note. Random effects=standard deviations; *p < 0.05.*

where Y_{ij} represents the outcome variable (e.g., time scientifically reasoning). λ_{0j} and λ_{1j}, respectively, represented the average initial status and average growth rate for predictors. Symbols σ_{0j}, σ_{1j}, and e_{ij}, represent, respectively, residual variance at initial status, residual variance in growth rate, as well as within-individual residual variance.

6 Results

Prior to calculating the analyses, descriptive statistics were estimated across the predictor and outcome variables to gauge the extent of time learners engaged in all and different scientific-reasoning actions as well as time facially expressing frustrated, confused, and neutral states across the agency conditions (Table 2).

6.1 To What Extent Does Time Scientifically Reasoning Change Over Time?

To describe and partition the variance of time scientifically reasoning during game-based learning, we built an unconditional means model to estimate the trend of time scientifically reasoning during game-based learning with Crystal Island between experimental conditions. The coefficients of the unconditional means model are provided in Table 3. Specifically, the variance components of the unconditional means model showed that approximately 1% of the variance in scientifically reasoning was explained between individuals, where the average time scientifically reasoning was 31.48 seconds and statistically significant from zero ($p < 0.05$). The 95% confidence interval of average time scientifically reasoning was between 29.71 to 33.32 seconds at each instance.

Next, we added a linear slope to the Null model to assess whether adding game play time (i.e., unconditional growth model) assisted in explaining variation in the time scientifically reasoning. The results indicated that by adding game play time as a predictor, it explained more variation in time scientifically reasoning between (as compared to within) individuals from 1% to 2% (Growth model in Table 3). This indicates that total time a learner spent in the game played a role on time spent scientifically reasoning, where the average time was 49.55 seconds and significant from zero ($p < 0.05$; Table 3). The model suggested a negative growth rate, such that the more seconds spent learning with Crystal Island was related to less seconds scientifically reasoning (b=-0.01, 95% CI: [-0.009, 0.007], $p < 0.05$). Although, it is important to note this negative association was small since actions were captured in seconds.

To examine the best model fit, we modeled the slope as randomly varying to assess whether time scientifically reasoning was significantly varying within individuals. Results suggested a negative growth rate, and the model found that adding game time as a random slope increased the proportion of variance from 2% to 8% (Mixed Model in Table 3). The average time scientifically reasoning was 51.06 seconds and statistically significant from zero ($p < 0.05$), where the more seconds in game play was related to a decrease in seconds scientifically reasoning (b=-0.01, 95% CI: [-0.009, -0.007], $p < 0.05$). Additionally, the variance in initial status of game time in Mixed Model suggested that some learners engaged in actions for longer periods at the start of the game relative to other learners. The growth rate was also significant ($p < 0.05$), showing there were individual differences in the rate of growth for game play time. However, there were no differences in time scientifically reasoning between experimental conditions ($p > 0.05$). This partially supported our hypothesis for research question 1, showing that seconds scientifically reasoning changed over the course of game-based learning, but that agency did not play a role in time scientifically reasoning with Crystal Island.

6.2 To What Extent are There Relationships Between Time Expressing Frustration, Confusion, and Neutral States and Time Scientifically Reasoning

To estimate whether and to what extent initial levels and dynamics of time facially expressing frustration, confusion, and neutral affective states predicted the trajectories of time scientifically reasoning, we calculated growth curve models. Specifically, we built three separate models for each facial expression of affective state variables: frustration, confusion, and neutral.

TABLE 4
Multilevel Growth Estimates of the Effects of Time Facially Expressing Frustration, Confusion, and Neutral Emotions on Time Scientifically Reasoning Over time

	Frustration	Confusion	Neutral
Variable	Estimate (*SE*)	Estimate (*SE*)	Estimate (*SE*)
Fixed effects			
Mean initial status	114.97 (27.13)	**140.38*** (21.11)	**92.59*** (29.43)
Mean growth rate	**-0.05*** (0.01)	**-0.05*** (0.01)	**-0.05*** (0.01)
Emotion duration	**0.32*** (0.05)	**0.26*** (0.03)	**0.40*** (0.05)
$Action_1$	**-115.30*** (4.54)	**-116.37*** (3.77)	**-114.38*** (1.93)
$Action_2$	**-84.19*** (2.93)	**-84.78*** (4.11)	**-83.57*** (2.25)
$Action_3$	**-116.41*** (12.58)	**-117.03*** (7.03)	**-114.84*** (5.35)
Pre-test scores [Level 2]	-65.89 (50.04)	**-93.00*** (38.86)	-53.84 (53.39)
Experimental $condition_1$ [Level 2]	**50.31*** (15.26)	**33.03*** (12.03)	-4.94 (16.28)
Random effects			
Within-individual	53.41	53.55	53.31
Initial status	82.31	62.89	97.66
Growth rate	**0.05***	0.04	**0.05***
Emotion duration	**0.37***	**0.27***	**0.40***
$Action_1$	37.55	**30.18***	10.51
$Action_2$	22.10	**33.67***	14.53
$Action_3$	110.29	**60.68***	45.38
Deviance	172787	172672	172575
ICC	0.70	0.58	0.77
Pseudo-R^2 (fixed effects)	0.25	0.30	0.35
Pseudo-R^2 (total effects)	0.90	0.85	0.91

Note. *$p < 0.05$; $Action_1$=Information gathering, $Action_2$=Hypothesis generation; $Action_3$=Experimental testing.

6.2.1 Frustration

The first model estimated whether the stability (i.e., intercept) and affective dynamics (i.e., slope) in time facially expressing frustration was related to the trajectories of time scientific reasoning during game-based learning (see Table 4). The fitted model estimated that, on average, learners scientifically reasoned for approximately 114.97 seconds at the start of learning with Crystal Island, and that the average rate of growth for 1 unit increase in seconds scientifically reasoning was related to an average 0.5 decrease in growth rate of seconds scientifically reasoning over the course of game-based learning while holding other predictors constant. Specifically, there was also a positive association between time expressing frustrated states and scientific reasoning. This highlighted that, while holding all else constant, for each 1 second increase in facially expressing frustration, was associated with a 0.32 second increase in time scientifically reasoning (fixed effects). The model also estimated negative associations across all scientific-reasoning action types (i.e., information gathering, hypothesis generation, and experimental testing), highlighting that the longer time learners scientifically reasoned overall, the less frequently they engaged in that action (relative to nonscientific-reasoning actions such as moving to different buildings). Results also suggested significant, positive relationship between time scientifically reasoning and participants assigned to the restricted agency condition relative to the full agency. There were no relations between pre-test scores and time scientifically reasoning.

The variance components demonstrated significant variation within individuals for the growth rate of time scientifically reasoning and duration of facially expressing frustrated states. The model also demonstrated that adding the predictors (relative to the unconditional growth model) increased the proportion of ICC by 68% and that 65% of the variance explained in the dependent variable existed between individuals (random effects). The model showed that 90% of the variance was explained by both the fixed and random effects, leaving approximately 10% of the variation in seconds scientifically reasoning unexplained.

6.2.2 Confusion

To examine the extent to which the stability (i.e., intercept) and affective dynamics (i.e., slope) in time facially expressing confusion was related to time scientifically reasoning, a second growth curve model was estimated (Table 4). This model showed that, on average, learners scientifically reasoned for approximately 140.38 seconds at the start of the game, and that the average rate of change for a 1 second increase in time scientifically reasoning was associated with a 0.05 decrease in the growth rate of time scientifically reasoning over the course of game-based learning, while holding all other predictors constant. Specifically, there was also a positive association between time expressing confused states and scientific reasoning. This highlighted that, while holding all else constant, for each 1 second increase in expressing a confused state, was associated with a 0.26 second increase in scientific reasoning (fixed effects). The model also suggested negative relationships between time scientifically reasoning and action type: information gathering, hypothesis generation, and experimental testing, respectively (fixed effects). Results also showed significant relationships between time scientifically reasoning and Level-2 predictors: pre-test scores and experimental condition, but only for participants assigned to the restricted agency condition relative to the full agency. This suggests that agency might have a positive moderating affect on time scientifically reasoning.

Further, the variance components suggested that the growth rate of time scientifically reasoning and the duration of time facially expressing confused states significantly varied within individuals (*ps* < 0.05), suggesting that individual differences may play a role in relations between seconds scientifically reasoning and the duration of confusion during game-based learning. The fitted model estimated significant, positive randomly effects for action types on time scientifically reasoning. This model demonstrated that adding the predictors increased the proportion of ICC by 56% and that 55% of the variance explained in the dependent variable existed between individuals. More specifically, the model showed that 85% of the variance was explained by both the fixed and random effects, leaving approximately 15% of the variation in seconds scientifically reasoning unexplained.

6.2.3 Neutral

Last, we estimated a third growth curve model to examine the extent to which stability (i.e., intercept) and affective dynamics (i.e., slope) in time expressing a neutral state was related to time scientifically reasoning during game-based learning (Table 4). This model showed that, on average, learners scientifically reasoned for approximately 92.59 seconds at the start of the game, and that the average growth rate for 1 second increase in scientific reasoning overall, there was a 0.05 second decrease in seconds scientifically reasoning over the course of game-based learning, while holding all other predictors constant. There was also a significant, positive relationship between time expressing a neutral state and the time scientifically reasoning. While holding all else constant, for a 1 second increase in expressing a neutral state, there was a 0.40 second increase in scientifically reasoning (fixed effects). This showed that the longer a neutral state was expressed, it was associated with longer engagements in scientific reasoning across game play. The model also suggested negative relations between time scientifically reasoning and action type: information gathering, hypothesis generation, and experimental testing, respectively (fixed effects). Results did not find significant relationships between time scientifically reasoning and Level-2 predictors: pre-test scores and experimental condition (*ps* > 0.05).

Further, the variance components indicated that the growth rate of time scientifically reasoning significantly varied within individuals ($p < 0.05$), showing individual differences (i.e., within-subjects variation) explained the initial level (i.e., intercept) and affective dynamics (i.e., slope) of time engaging in actions during game-based learning. The model also found that neutral state duration significantly varied within individuals ($p < 0.05$), showing individual differences played a role in variation explained in time engaging in scientific reasoning. This model demonstrated that by adding the predictors, it increased the proportion of ICC by 75% and that 56% of the variance explained in the dependent variable existed between individuals. More specifically, the model showed that 91% of the variance was explained by both the fixed and random effects, leaving approximately 9% of the variation in seconds scientifically reasoning unexplained.

7 Discussion

Several studies suggest affect is associated with higher-order thinking like scientific reasoning with emerging technologies such as game-based learning environments. Yet, significant gaps in literature exist because few studies employ sophisticated methodological and analytical techniques built to define, capture, and analyze affective dynamics and its relation to higher-order thinking over time during game-based learning using granular time series multichannel data. To address these issues, we examined relationships between time expressing frustration, confusion, and neutral states and time scientifically reasoning.

7.1 Research Question 1

The first model showed that the average time scientifically reasoning was 51.06 seconds, and that the longer learners engaged with Crystal Island, the less time they scientifically reasoned. Although, variance in time scientifically reasoning was explained within learners, suggesting that some learners engaged in scientific reasoning for longer periods of time relative to other learners. Further, the findings showed that there were individual differences in the rate of growth in time scientifically reasoning across game-based learning. This supported our hypothesis for research question 1, showing that time scientifically reasoning changed across time. These findings are consistent with the scientific reasoning as ICALM framework [46] and previous research that suggests individual characteristics play a role in scientific-reasoning behaviors across different emerging technologies [45], [51].

7.2 Research Question 2

Next, we examined associations between affective dynamics of expressing frustration, confusion, and neutral states and the trajectories of time scientifically reasoning with Crystal Island. Our models estimated that the average growth rate of time scientifically reasoning was negatively associated with time engaging in scientific reasoning, suggesting that the average growth rate of time scientifically reasoning was related to less time engaging in scientific reasoning over the course of game-based learning. Additionally, across the multilevel models, there were significant negative relations between the trajectory of time scientifically reasoning and action type, meaning that increases in the trajectory of scientific reasoning were associated with decreases in time engaging in specific scientific-reasoning actions over time. Further, we found significant relations between experimental condition, where learners assigned to the restricted agency condition showed positive associations with time engaging in scientific reasoning compared to the full agency conditions, particularly for multilevel growth models with frustrated and confused predictors compared to neutral predictors. A possible explanation for this finding could be that agency plays a moderating role on relations between affect (e.g., frustrated and confused states) and the trajectory of time scientifically reasoning during game-based learning. Specifically, when agency is restricted in pursuit of fostering scientific reasoning about relevant content and materials, it may have a positive moderating effect on relations between confusion, frustration, and scientific reasoning. However, more research is needed to adequately support this hypothesis. Future studies should pay closer attention to the degree of autonomy that learners are granted during game-based learning to assess its potential confounding influence on affect, its dynamics, and cognition [12].

The fitted model estimated significant positive relations between the duration of frustrated, confused, and neutral states and the trajectory of scientific reasoning during game-based learning. This finding supported our hypothesis, where we expected relationships to exist between affective states and the trajectory of scientific reasoning during game-based learning. The results also suggested that time expressing frustration, confusion, and neutral states significantly varied within individuals, and explained a great deal of variation in the trajectory of scientifically reasoning across (see Table 4). This was interesting and a possible explanation could be that individual characteristics play a moderating role on relationships between affective states and higher-order cognition. This finding is partially consistent with the ICALM model [7] which suggests that negative affect reduces cognitive resources needed to initiate and engage in higher-order thinking within working memory. A possible explanation for this finding could be that regulating confusion may have a different psychological basis than when regulating frustration and thus may be affecting working memory overload differently depending on the action being initiated or context (e.g., experimental testing versus generating hypotheses versus gathering information). There are additional gaps in research as few studies and contemporay theories (e.g., ICALM) because little to no studies explain the role of neutral states on cognitive resources or its role in affective dynamics. For example, do neutral states require cognitive resources to sustain? Does a neutral state indicate the effective use of an emotion regulation strategy, and to what extent does time expressing neutral states require cognitive resources (e.g., cognitively reappraising frustration in order to sustain a neutral state, etc.)? Previous research highlights mixed findings regarding relationships between confusion, frustration, and higher-order thinking [24], [26], [27], but little research has studied neutral states and their role on relations between affective dynamics and cognition with game-based learning environments. We suspect that relying on one data channel i.e., facial expressions of emotions to define, capture, and analyze associations between affective dynamics and higher-order cognition like scientific reasoning may not fully represent the complex nature of affective states. Including other data channels that represent the subsystems of affect would be a promising direction for future researchers to consider. For example, what might data channels on physiology, self-reported emotions, and emote-aloud protocols [54] offer more insight into how to fully represent affective states over time. We explain more in the future directions subsection below.

7.3 Future Directions

Studying affective states dynamically presents a promising avenue to study affective states and cognition during learning with emerging technologies across time and contexts. The temporal component of an affective experience provides a dimension of data that is rarely accounted for, and to gain more insight into this relationship, future research should utilize contemporary emotion regulation frameworks and multimodal data to detect potential detrimental affective states to assist individuals in effective emotion-regulation strategies such that they might be cognitively capable of making informed decisions driven by reason [52]. A start in this direction could be utilizing a multimodal mixed methods approach to capture the entire affect process [6], [39] like antecedents of an affective state of learners using both quantitative and qualitative data. What content were they viewing prior to the confused affective state, and could that pinpoint where the impasse stemmed from? Might it shed light on their goals and values and potentially help answer the question of why an affective state emerged in the first place to inform an effective emotion-regulation strategy if necessary? For example, if there was an impasse in goal pursuit (e.g., doing well in the course) and a learner expressed confusion while viewing the content, the system may be able to detect how to best assist the learner in regulating their affect by accounting for individual characteristics and then monitoring where and for how long they were fixating on content (e.g., was the content they were fixating on related to the goals of the learning session) and did they express confusion while reading about a particular topic? This data could inform ways to effectively provide scaffolding and pedagogical techniques to assist with emotion regulation into the design of technologies capable of monitoring and recognizing potentially detrimental affective dynamics that interfere with higher-order thinking [7], [24].

7.4 Limitations

In this study, most of our sample were Caucasian undergraduates across North America which does not represent most learners. Additionally, it is essential to acknowledge the limitations of facial recognition software that was trained using facial features captured from a largely white-male sample, potentially presenting limitations in recognizing facial expressions of affect from other cultures, races, and genders. Further, it is important to emphasize that when only collecting facial recognition data, there is an inherent assumption that facial expression displays are indicative of internal affective states. We acknowledge this limitation but observe that that facial recognition software offers continuous, granular data streams that can provide a mechanism for studying affective dynamics. Future researchers should collect streams of data on affective states like self-reports converged with facial expressions, physiology, and retrospective interviews to validate an affective state presence.

8 Conclusion

Evidence suggests affect is associated with higher-order thinking about complex topics with game-based learning environments [1]. Studies find mixed results regarding relationships between negative affect and higher-order thinking [26], [27], [28], [32], [53]. We argue these mixed findings result from few studies implementing advanced methodological and analytical techniques to align with contemporary theories describing affective dynamics and their relationship with higher-order thinking using granular time series data from multiple sensors. Our findings provide implications for addressing an essential aspect of designing affect-aware technologies: accounting for the dynamics of affective experiences and their relation to scientific reasoning. Advancing our understanding of the complex role of affect as it relates to higher-order thinking could provide a framework for designing affect-aware technologies that fuse multimodal data to captures the entire affective experience including its dynamics [4]. If a system can detect detrimental affective experiences (e.g., confusion persisting for too long), we have opportunity to design intelligent features within these systems to assist individuals in regulating their negative affective, such as prompting effective emotion-regulation strategies [18] to enhance their capacity to process information and utilize higher-order thinking skills across time.

ACKNOWLEDGMENTS

We would like to thank our collaborators in the Center for Educational Informatics at North Carolina State University. The authors would like to thank Robert Taylor, Michelle Taub, Nicholas Mudrick, and Megan Price for their assistance with data collection and pipeline development. Any conclusions expressed in this material do not necessarily reflect the views of SSHRC.

REFERENCES

[1] J. So, C. Achar, D. Han, N. Agrawal, A. Duhachek, and D. Maheswaran, "The psychology of appraisal: Specific emotions and decision-making," *J. Consum. Psychol.*, vol. 25, no. 3, pp. 359–371, Jul. 2015.

[2] E. B. Cloude, F. Wortha, D. A. Dever, and R. Azevedo, "Negative emotional dynamics shape cognition and performance with MetaTutor: Toward building affect-aware systems," in *Proc. 9th Int. Conf. Affect. Comput. Intell. Interact.*, 2021, pp. 1–8.

[3] P. V. Rouast, M. T. P. Adam, and R. Chiong, "Deep learning for human affect recognition: Insights and new developments," *IEEE Trans. Affect. Comput.*, vol. 12, no. 2, pp. 524–543, Apr.–Jun. 2021.

[4] S. K. D'mello and J. Kory, "A review and meta-analysis of multimodal affect detection systems," *ACM Comput. Surv.*, vol. 47, no. 3, pp. 1–36, 2015.

[5] J. M. Harley, S. P. Lajoie, C. Frasson, and N. C. Hall, "Developing emotion-aware, advanced learning technologies: A taxonomy of approaches and features," *Int. J. Artif. Intell. Edu.*, vol. 27, no. 2, pp. 268–297, Jun. 2017.

[6] K. R. Scherer and A. Moors, "The emotion process: Event appraisal and component differentiation," *Annu. Rev. Psychol.*, vol. 70, pp. 719–745, 2019.

[7] J. L. Plass and U. Kaplan, "Emotional design in digital media for learning," in *Emotions, Technology, Design, and Learning*, Amsterdam, The Netherlands: Elsevier, 2016, pp. 131–161.

[8] K. Loderer, R. Pekrun, and J. C. Lester, "Beyond cold technology: A systematic review and meta-analysis on emotions in technology-based learning environments," *Learn. Instruct.*, vol. 70, 2020, Art. no. 101162.

[9] D. B. Clark, E. E. Tanner-Smith, and S. S. Killingsworth, "Digital games, design, and learning: A systematic review and meta-analysis," *Rev. Educ. Res.*, vol. 86, no. 1, pp. 79–122, 2016.

[10] P. Wouters, C. van Nimwegen, H. van Oostendorp, and E. D. van der Spek, "A meta-analysis of the cognitive and motivational effects of serious games," *J. Educ. Psychol.*, vol. 105, no. 2, 2013, Art. no. 249.

[11] J. L. Plass, R. E. Mayer, and B. D. Homer, *Handbook of Game-Based Learning*. Cambridge, MA, USA: MIT Press, 2020.

[12] M. Taub, R. Sawyer, A. Smith, J. Rowe, R. Azevedo, and J. Lester, "The agency effect: The impact of student agency on learning, emotions, and problem-solving behaviors in a game-based learning environment," *Comput. Edu.*, vol. 147, 2020, Art. no. 103781.

[13] N. Cowan, "The magical number 4 in short-term memory: A reconsideration of mental storage capacity," *Behav. Brain Sci.*, vol. 24, no. 1, pp. 87–114, 2001.

[14] J. L. Plass and S. Kalyuga, "Four ways of considering emotion in cognitive load theory," *Educ. Psychol. Rev.*, vol. 31, pp. 339–359, 2019.

[15] L. Linnenbrink-Garcia, E. A. Patall, and R. Pekrun, "Adaptive motivation and emotion in education: Research and principles for instructional design," *Policy Insights Behav. Brain Sci.*, vol. 3, no. 2, pp. 228–236, 2016.

[16] R. Moreno, "Cognitive load theory: More food for thought," *Instructional Sci.*, vol. 38, no. 2, pp. 135–141, 2010.

[17] B. Park, L. Knörzer, J. L. Plass, and R. Brünken, "Emotional design and positive emotions in multimedia learning: An eyetracking study on the use of anthropomorphisms," *Comput. Edu.*, vol. 86, pp. 30–42, 2015.

[18] R. Pekrun and K. Loderer, *Emotions and Learning From Multiple Representations and Perspectives*. Evanston, IL, USA: Routledge, 2020.

[19] S. D'Mello, A. Kappas, and J. Gratch, "The affective computing approach to affect measurement," *Emotion Rev.*, vol. 10, no. 2, pp. 174–183, 2018.

[20] R. Pekrun, S. Lichtenfeld, H. W. Marsh, K. Murayama, and T. Goetz, "Achievement emotions and academic performance: Longitudinal models of reciprocal effects," *Child Develop.*, vol. 88, no. 5, pp. 1653–1670, 2017.

[21] M. J. Price, N. V. Mudrick, M. Taub, and R. Azevedo, "The role of negative emotions and emotion regulation on self-regulated learning with metatutor," in *Proc. Int. Conf. Intell. Tutoring Syst.*, 2018, pp. 170–179.

[22] S. P. Pattar, E. Coronado, L. R. Ardila, and G. Venture, "Intention and engagement recognition for personalized human-robot interaction, an integrated and deep learning approach," in *Proc. IEEE 4th Int. Conf. Adv. Robot. Mechatronics*, 2019, pp. 93–98.

[23] J. A. Russell, "A circumplex model of affect," *J. Pers. Social Psychol.*, vol. 39, no. 6, 1980, Art. no. 1161.

[24] S. D'Mello and A. Graesser, "Dynamics of affective states during complex learning," *Learn. Instruct.*, vol. 22, no. 2, pp. 145–157, 2012.

[25] R. S. Baker, J. L. Ocumpaugh, and J. Andres, "BROMP quantitative field observations: A review," in *Learning Science: Theory, Research, and Practice*. New York, NY, USA: McGraw-Hill, 2018.

[26] R. S. Baker, S. K. D'Mello, M. M. T. Rodrigo, and A. C. Graesser, "Better to be frustrated than bored: The incidence, persistence, and impact of learners' cognitive–affective states during interactions with three different computer-based learning environments," *Int. J. Hum.-Comput. Stud.*, vol. 68, no. 4, pp. 223–241, Apr. 2010.

[27] E. B. Cloude, F. Wortha, D. A. Dever, and R. Azevedo, "How do emotions change during learning with an intelligent tutoring system? Metacognitive monitoring and performance with MetaTutor," in *Proc. 42nd Annu. Conf. Cogn. Sci. Soc.*, 2020, pp. 423–429.

[28] W. Ahmed, G. van der Werf, H. Kuyper, and A. Minnaert, "Emotions, self-regulated learning, and achievement in mathematics: A growth curve analysis," *J. Educ. Psychol.*, vol. 105, no. 1, 2013, Art. no. 150.

[29] J. M. A. L. Andres et al., "Affect sequences and learning in Betty's brain," in *Proc. 9th Int. Conf. Learn. Analytics Knowl.*, 2019, pp. 383–390.

[30] P. Ekman, W. V. Friesen, and J. C. Hager, *The Facial Action Coding System: A Technique for the Measurement of Facial Movement*, Salt Lake City, Utah: A Human Face Salt Lake City, 2002, pp. 77–254.

[31] iMotions, "Attention tool," Version 6.3, Comput. Softw., iMotions, Inc., Boston, MA, USA, 2017.

[32] M. Taub, R. Sawyer, J. Lester, and R. Azevedo, "The impact of contextualized emotions on self-regulated learning and scientific reasoning during learning with a game-based learning environment," *Int. J. Artif. Intell. Edu.*, vol. 30, no. 1, pp. 97–120, 2020.

[33] F. Gabriel, E. Cloude, and R. Azevedo, "Using learning analytics to measure motivational and affective processes during self-regulated learning with advanced learning technologies," in *Social and Emotional Learning and Complex Skills Assessment*, Cham, Switzerland: Springer, 2022.

[34] S. W. McQuiggan, J. Goth, E. Ha, J. P. Rowe, and J. C. Lester, "Student note-taking in narrative-centered learning environments: Individual differences and learning effects," in *Proc. Int. Conf. Intell. Tutoring Syst.*, 2008, pp. 510–519.

[35] F. E. Grubbs, "Procedures for detecting outlying observations in samples," *Technometrics*, vol. 11, no. 1, pp. 1–21, 1969.

[36] P. Dente, D. Küster, L. Skora, and E. Krumhuber, "Measures and metrics for automatic emotion classification via facet," in *Proc. Conf. Study Artif. Intell. Simul. Behav.*, 2017, pp. 160–163.
[37] M. Taub, R. Azevedo, R. Rajendran, E. B. Cloude, G. Biswas, and M. J. Price, "How are students' emotions related to the accuracy of cognitive and metacognitive processes during learning with an intelligent tutoring system?," *Learn. Instruct.*, vol. 72, 2021, Art. no. 101200.
[38] S. Stöckli, M. Schulte-Mecklenbeck, S. Borer, and A. C. Samson, "Facial expression analysis with AFFDEX and FACET: A validation study," *Behav. Res. Methods*, vol. 50, no. 4, pp. 1446–1460, 2018.
[39] L. F. Barrett, R. Adolphs, S. Marsella, A. M. Martinez, and S. D. Pollak, "Emotional expressions reconsidered: Challenges to inferring emotion from human facial movements," *Psychol. Sci. Public Int.*, vol. 20, no. 1, pp. 1–68, 2019.
[40] "SMI Experiment Center," Version 3.4.165, Apparatus Softw., SensoMotoric Instruments, Boston, MA, USA, 2014.
[41] D. D. Salvucci and J. H. Goldberg, "Identifying fixations and saccades in eye-tracking protocols," in *Proc. Symp. Eye Tracking Res. Appl.*, 2000, pp. 71–78.
[42] G. van Rossum and F. L. Drake Jr, *Python Tutorial*, Amsterdam, The Netherlands: Centrum voor Wiskunde en Informatica, 1995.
[43] R. Team, *RStudio: Integrated Development Environment for R*, RStudio, PBC, Boston, MA, USA, 2020. [Online]. Available: http://www.rstudio.com/
[44] D. Bates, M. Mächler, B. Bolker, and S. Walker, "Fitting linear mixed-effects models using lme4," *J. Statist. Softw.*, vol. 67, pp. 1–48, 2015.
[45] E. B. Cloude, D. A. Dever, M. D. Wiedbusch, and R. Azevedo, "Quantifying scientific thinking using multichannel data with crystal Island: Implications for individualized game-learning analytics," *Front. Edu.*, vol. 5, pp. 217–238, 2020.
[46] D. Klahr and K. Dunbar, "Dual space search during scientific reasoning," *Cogn. Sci.*, vol. 12, no. 1, pp. 1–48, 1988.
[47] E. Alemdag and K. Cagiltay, "A systematic review of eye tracking research on multimedia learning," *Comput. Edu.*, vol. 125, pp. 413–428, 2018.
[48] J. Singer and J. Willet, *A Framework for Investigating Change Over Time*. New York, NY, USA: Oxford Univ. Press, 2003.
[49] T. A. Snijders et al., *An Introduction to Basic and Advanced Multilevel Modeling*. Newbury Park, CA, USA: Sage, 1999, pp. 13–524.
[50] S. G. Luke, "Evaluating significance in linear mixed-effects models in R," *Behav. Res. Methods*, vol. 49, no. 4, pp. 1494–1502, 2017.
[51] A. W. Lazonder and R. Harmsen, "Meta-analysis of inquiry-based learning: Effects of guidance," *Rev. Educ. Res.*, vol. 86, no. 3, pp. 681–718, 2016.
[52] K. McRae and J. J. Gross, "Emotion regulation," *Emotion*, vol. 20, no. 1, pp. 1–9, 2020.
[53] S. D'Mello, B. Lehman, R. Pekrun, and A. Graesser, "Confusion can be beneficial for learning," *Learn. Instruct.*, vol. 29, pp. 153–170, 2014.
[54] S. K. D'Mello, S. D. Craig, and A. C. Graesser, "Multimethod assessment of affective experience and expression during deep learning," *Int. J. Learn. Technol.*, vol. 4, no. 3–4, pp. 165–187, 2009.

Elizabeth B. Cloude is a postdoctoral research fellow within the Penn Center of Learning Analytics, University of Pennsylvania. Her research focuses on understanding dynamic and interdependent relations between cognitive, affective, metacognitive, and motivational processes that emerge during self-regulated learning with emerging technologies using multimodal learning analytics at both small- and large-scale. Her overall goals are to build personalized and adaptive educational technologies that support all types of learning needs using instructional design principles that guide effective teaching and learning practices.

Daryn A. Dever is currently working toward the PhD degree in the School of Modeling, Simulation, and Training, University of Central Florida. Her research is focused on studying learners' interactions with game- and hypermedia-based learning environments using multimodal multichannel data (e.g., eye tracking, log files, facial expressions of emotions) to inform design principles for personalized and adaptive learning environments.

Debbie L. Hahs-Vaughn is a professor with the Methodology, Measurement, and Analysis Program, College of Community Innovation and Education, University of Central Florida. Her primary research interest relates to methodological issues associated with applying quantitative statistical methods to survey data obtained under complex sampling designs and using complex survey data to answer substantive research questions.

Andrew J. Emerson received the bachelor's degree in computer science and mathematics from Furman University, and the PhD degree in computer science from North Carolina State University. He is a machine learning engineer with Educational Testing Service, where he works on the Multimodal AI Team. His research interests span predictive student modeling, multimodal learning analytics, and educational applications of machine learning.

Roger Azevedo is a professor with the School of Modeling, Simulation, and Training, University of Central Florida. His main research area includes examining the role of cognitive, metacognitive, affective, and motivational regulatory during human-machine interactions with advanced learning technologies. He has published more than 200 peer-reviewed papers in educational, learning, cognitive, and computational sciences. He is the former editor of the *Metacognition and Learning* journal. He is a fellow of the American Psychological Association.

James Lester is distinguished University professor of computer science with North Carolina State University, where he is director of the Center for Educational Informatics. His research on artificial intelligence-driven technologies for education ranges from adaptive learning environments and multimodal learning analytics to intelligent game-based learning environments, computational models of interactive narrative, and affective computing. He is the author of more than two hundred publications in AI technologies for education. He has served as editor-in-chief of the *International Journal of Artificial Intelligence in Education* and is a fellow of the Association for the Advancement of Artificial Intelligence (AAAI).