Emotions in Action: How Students' Regulatory Responses Shape Learning

Andres Felipe Zambrano
University of Pennsylvania
Philadelphia, Pennsylvania, USA
azamb13@upenn.edu

Jaclyn Ocumpaugh
University of Houston
Houston, Texas, USA
jlocumpaugh@gmail.com

Ryan S. Baker
Adelaide University
Adelaide, South Australia, Australia
ryanshaunbaker@gmail.com

Jessica Vandenberg
North Carolina State University
Raleigh, North Carolina, USA
jvanden2@ncsu.edu

Abstract

Emotions play a central role in shaping learning within digital environments. Although their effects may depend on how students' emotional experiences manifest into concrete behaviors, the links between these dimensions remain underexplored. This study investigates the most common behaviors during episodes of boredom, confusion, frustration, and engaged concentration in an educational game, as well as associations with situational interest, self-efficacy, prior knowledge, and learning gains, using interaction logs and sensor-free affect detectors. Results show that boredom is linked to off-task roaming, both consistently associated with lower motivation and learning. In contrast, behaviors during engaged concentration, frustration, and especially confusion vary widely, shaped by motivational traits and prior knowledge and offering diverse associations with learning. Concrete regulatory responses in these states—such as systematizing findings with in-game tools, skimming domain content to resolve doubts, or testing hypotheses—are positively associated with learning and motivation, reflecting students' ability to regulate emotions and address cognitive challenges. However, less constructive responses, such as aimless wandering, were tied to lower knowledge and motivation, underscoring the need for additional support. These findings extend existing affective theory by underscoring the importance of considering the behavioral dimension when analyzing students' emotions in digital learning environments.

CCS Concepts

• **Applied computing → Interactive learning environments**; • **Human-centered computing** → *Human computer interaction (HCI).*

Keywords

Educational Games, Affective States, Engagement, Emotion Regulation, Motivation

LAK 2026, Bergen, Norway

ACM ISBN 979-8-4007-2066-6/2026/04
https://doi.org/10.1145/3785022.3785039

ACM Reference Format:
Andres Felipe Zambrano, Jaclyn Ocumpaugh, Ryan S. Baker, and Jessica Vandenberg. 2026. Emotions in Action: How Students' Regulatory Responses Shape Learning. In *LAK26: 16th International Learning Analytics and Knowledge Conference (LAK 2026), April 27-May 01, 2026, Bergen, Norway*. ACM, New York, NY, USA, 12 pages. https://doi.org/10.1145/3785022.3785039

1 Introduction

Emotions in digital learning environments have been extensively studied within the learning analytics community, where there has been considerable effort to understand how students' affective experiences interact with cognitive, metacognitive, and motivational processes [19, 28]. Research consistently shows that low-arousal, negatively valenced emotions (e.g., boredom) hinder learning across these dimensions, whereas positively valenced emotions (e.g., engaged concentration) generally support learning outcomes [28]. However, high-arousal negative emotions (particularly confusion and frustration) do not exhibit such straightforward relationships with learning.

Several frameworks have sought to explain how these affective experiences relate to learning outcomes [15, 38, 43], with increasing attention being paid to possible differences in the forms affective states may take. For example, researchers have suggested that differences between canonical and pleasant frustration [22] or between several different kinds of confusion and frustration [2, 38] may arise from different causes and leading to varied impacts on learning. These frameworks seek to explain the complex relationships between affect and learning outcomes from motivational and cognitive perspectives, but they are not always concretely linked to the behavioral dimensions through which these emotions ultimately shape learning.

Some studies have looked at the interplay between affect and behavior, showing that boredom often precedes or motivates disengaged behaviors such as gaming the system or going off task [5], while engaged concentration tends to mitigate these behaviors while fostering more constructive ones [4, 21]. However, the range of behaviors students adopt during episodes of specific affective states, and the factors shaping which behavior a student chooses as a response to specific affect remain underexplored, particularly for emotions like confusion and frustration.

To expand understanding in this area, the present study examines the specific moments when students experience confusion, frustration, boredom, and engaged concentration, within the context of an educational game used by middle school students. We ask three research questions: (1) which behaviors are most common during episodes of each affective state; (2) which behaviors are associated with different motivational and learning measures; and (3) how these relationships shed light on the gaps on existing theory around affect during learning. By analyzing the actions students select during specific affective episodes and linking them to both learning outcomes and motivational measures, we seek to provide a more nuanced account of why different students choose different actions in response to the same emotion, and how emotions shape learning through the behaviors they elicit. A better understanding of these complex relationships can help fill in some of the gaps within current affective theory, with practical implications for educational design and adaptive learning systems.

2 Related Work

2.1 Theoretical Perspectives of Affect in Digital Learning Environments

Several frameworks have sought to explain how affective states arise and evolve in learning environments. One of the earliest is Csikszentmihalyi's [15] Flow Theory, which posits that flow, a positive state of immersion in a task, emerges when task difficulty and learner skill align and fosters autonomy, and sustained concentration, while insufficient challenge leads to boredom and excessive challenge leads to anxiety. Pekrun [43] extends this framework in discussions of Control Value Theory (CVT), suggesting that flow depends not only on the balance of skill and difficulty but also on perceived relevance. In his empirically based model, Pekrun argues that boredom may occur in both overly simple and overly difficult tasks, but it may also occur when a student is asked to do something that is matched to their skill but is not perceived as valuable by the student.

Building on these perspectives, D'Mello and Graesser [20] proposed a dynamic model of affect in computer-based learning, emphasizing transitions among engaged concentration, confusion, frustration, and boredom. In this model, confusion (cognitive disequilibrium) arises when learners encounter an impasse; this may catalyze deeper learning if resolved, or escalate into frustration and eventually boredom if left unresolved. The model positions engaged concentration as generally positive and boredom as consistently negative, while treating confusion and frustration as ambivalent states—potentially constructive or detrimental depending on whether they transition back to engagement or toward disengagement.

Most recently, Ocumpaugh et al. [38] introduced the Skills, Difficulty, Value, Efficacy, and Time (SDVET) model, which integrates cognitive and motivational constructs to explain when and why learners transition between affective states. Like Flow Theory, SDVET predicts boredom when skills exceed task difficulty, but it also acknowledges that frustration can occur when skills fall below task demands. Unlike Flow Theory, it does not assume that the space between these extremes is fully occupied by flow or engagement. Instead, drawing on CVT, it argues that even when skill and difficulty are balanced, canonical "unpleasant" frustration may still arise if learners lack perceived value, whereas valued and interesting tasks are more likely to elicit flow. When students encounter tasks above their current skill level but still perceive them as meaningful and manageable (high value and self-efficacy), they are more likely to experience "pleasant" frustration [22], choosing to persist rather than withdraw. However, when task value or self-efficacy is low, students may experience "intolerable" confusion or frustration, which can devolve into boredom, even without an apparent cognitive impasse. The SDVET model also suggests that the buffering effect of value and self-efficacy is time-limited. Even highly self-efficacious students may struggle to sustain pleasant frustration as either novelty or interest fades. The cognitive and emotional burden of prolonged self-regulation can erode this capacity and eventually lead to disengagement.

Recognizing similarities between certain forms of both confusion and frustration, Baker et al. [2] propose reframing them not as separate, internally uniform states but as part of a "confrustion constellation" of interrelated affective subtypes that share common features yet differ in their triggers and effects. Confusion and frustration often overlap because learners may experience them sequentially or simultaneously. However, some forms of confusion and frustration can also diverge substantially. For instance, Gee's [22] notion of pleasant frustration aligns more closely with brief episodes of confusion that are quickly resolved and keep the student engaged rather than with deep, unpleasant frustration stemming from low value or motivation. Baker et al. [2] therefore argue that some forms of frustration are closer to certain forms of confusion than to other types of frustration (and vice versa). This perspective supports viewing confusion and frustration as a constellation of related subtypes rather than as two rigidly distinct affective states.

2.2 Associations between Affect and Learning in Digital Learning Environments

Empirical findings across a wide range of digital learning environments provide strong support for many aspects of the frameworks described above. For example, Karumbaiah et al. [28] conducted a comprehensive review of the impacts of confusion, frustration, engaged concentration, and boredom across multiple contexts. They found consistent evidence of a positive association between engaged concentration and a variety of outcomes, including performance on knowledge tests (e.g., [12]), learning gains (e.g., [21, 27]), standardized state tests (e.g., [30, 41]), and even long-term indicators such as college enrollment years later [51]. While a small number of studies report null findings (e.g., [23]), none have identified negative associations between this affective state and outcomes. In addition, motivational factors such as initial self-efficacy and situational interest have been linked to longer and more frequent episodes of engaged concentration and more transitions into this state [55–57], which in turn are associated with learning gains and higher motivational scores on post-tests [36, 38]. This evidence consistently suggests that positive motivational factors, such as situational interest and self-efficacy, enhance the likelihood of experiencing engaged concentration, which in turn drives positive learning outcomes.

Similarly consistent (but in the opposite direction) is the evidence that boredom is detrimental to learning. Multiple studies in Karumbaiah et al.'s review show negative associations between boredom and learning gains [9, 23], exam scores [30, 47], and college enrollment [51]. Although some studies have found null effects [24] or in one unusual case a slightly positive effect (seen in two different learning systems [21, 41]), there is no strong evidence of boredom having positive impacts on learning. Further research on motivational factors has found that low situational interest or self-efficacy can act as precursors to boredom, in turn reinforcing lower interest and engagement in learning [38, 54, 55].

Although the associations are relatively consistent for boredom and engaged concentration, much less is clear about frustration and confusion. Karumbaiah et al. [28] highlighted strong disagreements in the literature regarding these two states. For confusion, many studies have reported null effects for cognitive [17, 18] and motivational measures [55], while others have found positive associations with learning [21, 24], particularly when confusion is brief and successfully resolved [32]. However, some studies have documented negative associations between confusion and exam performance [30, 47], long-term outcomes such as college enrollment [51], and particularly longer episodes of confusion for students with lower situational interest and self-efficacy [56]. These results suggest the existence of a zone of optimal confusion in which confusion has the potential to support deeper understanding when learners are able to resolve their confusion, whereas confusion that persists without resolution may become detrimental and contribute to disengagement [1].

Similarly, most studies on frustration report null findings (e.g., [17, 24, 30]), although some have identified negative associations with within-platform performance [9, 40]. However, in one game context, frustration was positively associated with performance and learning [53]. Additionally, students with higher self-efficacy are more likely to experience longer episodes of frustration that eventually transition into positive affect (e.g., concentration or delight; [38, 56]) during gameplay. This pattern suggests that frustration, under certain conditions and mainly for educational games, can contribute to more positive and motivating learning experiences, supporting the existence of Gee's theorized pleasant frustration [22].

Even when confusion and frustration are considered together, evidence is similarly mixed. Richey et al. [45, 46] found that although confrustion was negatively associated with post-test and delayed post-test scores overall, it tended to occur more frequently when students learned from erroneous examples, which in turn supported better test performance. Likewise, Liu et al. [34] reported that brief episodes combining confusion and frustration were positively related to learning. Thus, while the effects of boredom and concentration are consistent, the effects of confusion and frustration seem to be more conditional. They hinder learning in some contexts, but some mechanisms, such as self-efficacy or motivation, timely resolution of impasses, or supportive task design, can also render them constructive.

2.3 Behavioral Pathways of Emotions in Learning

Multiple studies have examined how affective states and behaviors interact to shape learning outcomes. For instance, Fancsali [21] found that boredom increases off-task behaviors, whereas engaged concentration reduces gaming the system (where students exploit platform features to generate correct answers without genuine engagement), which has been consistently linked to negative outcomes. These findings align with other work showing that on-task behaviors are most often associated with engaged concentration [4, 7, 48]. Although Fancsali [21] observed a small negative association between boredom and gaming the system overall, studies across other learning environments typically find that boredom precedes this behavior [4, 7, 48]. Similarly, Bosch and D'Mello [9] found that curiosity and engagement typically followed reading or coding, whereas confusion and frustration followed errors and preceded hint use. Of these emotions, only frustration was negatively associated with learning, while confusion had a positive association. These findings suggest that when confusion motivates help-seeking and is resolved before escalating, the outcome can be beneficial. However, when confusion persists and develops into frustration, learning outcomes are more likely to be negative.

Although frustration sometimes appears detrimental to learning, its role in educational games may be more complex, as pleasant frustration can arise from games' narrative and immersive features [22], making relationships less straightforward than in other learning contexts. For example, in the context of Crystal Island, an open-world scientific inquiry game [49], Cloude et al. [13] reported that both confusion and frustration were positively associated with time spent on science-related actions, suggesting that these emotions may sometimes support constructive behaviors. For the same game, Sabourin and Lester [50] used student self-reports to examine how different responses to affective episodes can either support or hinder learning. The authors found that students who stayed engaged in scientific tasks after experiencing confusion were more likely to report later engaged concentration and achieved greater learning than those who disengaged to explore the virtual world. In contrast, when students experienced frustration, persisting with scientific tasks was unhelpful, while taking a break by exploring the virtual world appeared to yield better learning outcomes. Because the self-reports in this study were not very granular, Sabourin and Lester [50] also applied machine learning-based detectors that could distinguish between positive and negative affect, also showing that taking a break during negative affect was beneficial.

More recently, Zambrano et al. [55] developed detectors for Crystal Island that can identify specific affective states. These detectors have been used to examine affective dynamics across behavioral archetypes, revealing that students engaged in scientific actions are more likely to experience confusion and frustration, whereas disengaged students who tend to roam are more likely to experience boredom [57]. Despite these advances, little is known about the specific actions students take during affective episodes, how these behaviors may depend on factors such as initial motivation or prior knowledge, and how they differentially impact multiple student outcomes.

3 Methods

3.1 Learning Platform

In this study, we analyzed data from 122 middle school students who interacted with Crystal Island as part of their regular science

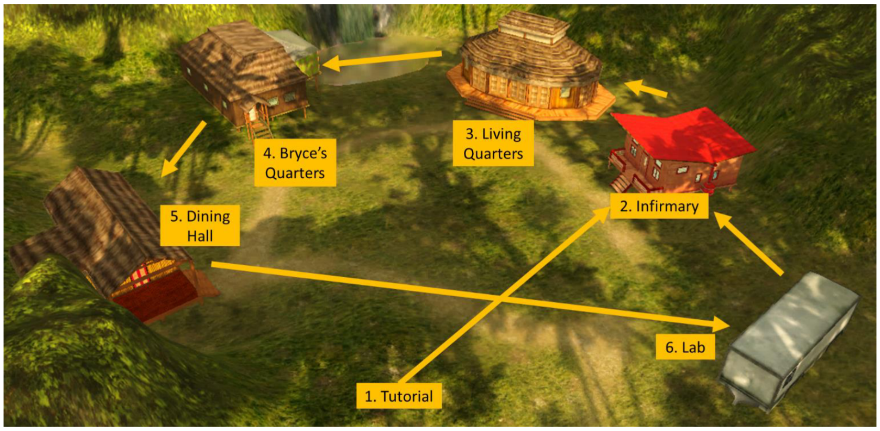

Figure 1: Overview of Crystal Island with the expert "golden pathway" for game completion as operationalized by [52].

curriculum. Crystal Island is a single-player, open-world game designed to spark interest and support inquiry-based learning in microbiology [49]. In the game, players assume the role of an investigative scientist working to diagnose a mysterious illness affecting a research station on a remote island. Their goal is to identify the disease-causing pathogen, how it is being transmitted, and what is the most suitable treatment to address the outbreak. Through gameplay, students build microbiology knowledge related to disease transmission, treatment, and prevention, while also developing scientific reasoning skills to analyze symptoms, identify carriers, and track patterns of infection.

The gameplay begins in a tutorial area where players are introduced to the fundamental mechanics of the game, such as interacting with non-player characters (NPCs) to gather information, collecting readings, submitting concept matrices (required at the end of each reading to assess understanding and can be submitted multiple times), picking up objects, and testing hypotheses. Upon completing the tutorial, players are directed to the infirmary to receive further instructions, although they are free to explore the island as they choose (see Figure 1). As players collect information, they are expected to gather objects across the island and scan them in the laboratory, formulating and testing hypotheses regarding the potential virus or bacteria causing the outbreak.

3.2 Data Context

The data used in this study was collected in 2023 at an urban middle school in the southeastern United States. The participant group was gender-balanced, with 53 students identifying as male, 66 as female, and 5 choosing to self-describe. The sample was demographically diverse, including strong representation from groups historically underrepresented in STEM: 46% of students identified as Black, 16% as Hispanic, 5% as Asian, 5% as Multiracial, and 1% as Native American.

Four members of the research team were present during the classroom implementation: two as non-interactive observers and two who introduced the activity and provided technical support as needed. At the start of the study, students completed a series of established, validated measures: a demographic questionnaire; a science content pre-test scored from 0 to 17; and surveys assessing self-efficacy [11], situational interest [33], and game literacy. The game literacy scale ranged from 0 (no weekly video game play) to 3 (more than 10 hours per week). These instruments were selected to assess students' baseline knowledge, interest, self-efficacy, and familiarity with games—factors hypothesized to influence their engagement with the learning experience.

Upon completing the initial surveys and pre-test, students played Crystal Island over two class periods across two days (approx. 60 min. of gameplay in total). During gameplay, students were asked to self-report their affective states at specific points (i.e., in-game milestones) to reduce disruption to their learning experience.

After playing the game, they completed a set of post-game surveys. These included a science content post-test (identical to the pre-test) to assess knowledge gains, along with five subscales from the Intrinsic Motivation Inventory (IMI; [16]): interest-enjoyment (IE), perceived competence (PC), effort-importance (EI), pressure-tension (PT), and value-utility (VU). Unlike the earlier self-efficacy and situational interest measures, which assessed general attitudes, these post-game subscales focused on students' perceptions of their experience with the game and were therefore only administered after gameplay. To reduce the risk of survey fatigue or adverse reactions from excessive questioning [44], the self-efficacy and situational interest surveys were not repeated. All post-game subscales were revalidated using Cronbach's alpha within this dataset (see [54]).

3.3 Analysis

Affective labels were generated by trained cross-validated, sensor-free affect detectors that estimate the probability of students experiencing each affective state (one vs. all detectors) in 20-second intervals [55]. The use of a 20-second window follows standard practice in affective research [6] and affective chronometry [10]. For the behavioral data, we adapted Zambrano et al.'s [54]) codebook, which was used to analyze student behaviors independently of affective data (see Table 1). Specifically, we adjusted the time scale of the codes *Long Time Outside* and *Inside No Action* to better align with the granularity of the affect detection and the duration

Table 1: Codebook (Adapted from [54]).

Behavior	Definition
Inside No Action	20 consecutive seconds inside any location without conducting any additional action (original definition by Zambrano et al. was 2 minutes).
Outside No Action	*Returning to Tutorial*: Return to Tutorial after visiting any other location *Excessive Time Outside*: 20 consecutive seconds outside (original definition by Zambrano et al. was 2 minutes).
Rushing	*Rushed Reading*: Less than 10 s of reading for each article. *Rushed Conversation*: Less than 5s interacting with an NPC (repeated conversations are not considered) or less than 1SD below the median time spent speaking with a specific NPC
Long Conversation	Over the median of time speaking with the specific NPC or more than 20 consecutive seconds.
Worksheet	Students add an element to the worksheet.
Scan After Exploration	*Scan After Reading*: Students read for at least 5 min before testing a hypothesis (at any point of the gameplay, not necessarily during the confusion or frustration episode). *Scan After Conversation*: Students had at least 3 min of conversations before testing a hypothesis (at any point of the gameplay, not necessarily during the confusion or frustration episode).
Repeated Testing	*Repeated Hypothesis Testing*: 3 consecutive hypothesis tests without adding data to the worksheet, reading again, or conversing with an NPC, with a separation of less than 20 s between each scan *Repeated Concept Matrix Testing*: 3 consecutive concept matrix submissions in less than 10 seconds. (Previously labeled *No Reflection* in [54]).

of affective episodes. Whereas Zambrano et al. applied these labels after two minutes of inactivity, we applied them at 20-second intervals to align with the affect labeling periodicity (clip duration). Because the original label *Long Time Outside* no longer implied extended inactivity when applied every 20 seconds, we renamed it *Outside No Action*. Similarly, Zambrano et al.'s code *No Reflection* was renamed *Repeated Testing*, which more accurately describes the observed behavior—repeatedly testing hypotheses or submitting concept matrices within a short period.

After labeling each affective episode with at least one corresponding behavior, we calculated the average frequency of each behavior across students' episodes of confusion, frustration, engaged concentration, and boredom according to the SR-based detectors. We used relative behavior frequencies per clip to control for differences in the typical duration of each affective state [56]. This adjustment ensured that longer-lasting states (e.g., boredom) did not disproportionately inflate behavior counts. We excluded nervousness and happiness from the analysis because they were relatively infrequent; over 40% of students never reported experiencing either of these states.

We next conducted a Friedman χ^2 test to assess differences in behavior frequency during confusion, frustration, boredom, and engaged concentration. This non-parametric test accounts for repeated measures from the same students (one observation per affective state) and enabled us to analyze how behavior varied across the four emotions. Next, we calculated Kendall's τ correlation coefficients between the behavior frequencies and five external measures: Situational Interest, Self-Efficacy, Interest Engagement, pre-test scores, and learning gains. These measures were selected because they most directly capture students' interest and learning outcomes before and after gameplay. We intentionally limited the number of external measures to reduce the risk of false discoveries. Although the zero-inflated distribution of behaviors was less pronounced in the SR affect data, it was still present for *Outside No Action*, *Rushing*, and *Repeated Testing* when split by affective state. As these distributions produced many tied ranks, we used Kendall's τ instead of Spearman's ρ [29].

We also assessed the monotonicity of all identified significant associations using generalized additive models (GAM), after applying a rank transformation to the data, to observe monotonic associations as linear. This analysis examined whether students who never displayed the behavior during episodes of the corresponding affect (i.e., occurrence=0) exhibited patterns that differed from those expected based on the estimated Kendall's τ, which could indicate non-linearity or other relationships beyond what a single correlation coefficient can capture. This additional step is particularly important given the zero-inflated distributions of less frequent behaviors (e.g., *Rushing*, *Repeated Testing*, and *Long Conversations*). The GAMs were fitted with polynomial splines, allowing the relationship to take the shape of a polynomial of up to degree 5, to model the association between variables.

Since we computed 35 correlations for each affective state (7 behaviors × 5 measures), we applied the Benjamini-Hochberg procedure to control the false discovery rate [8]. However, because this correction is overly conservative when applied to the total number of correlations per affect (i.e., it would require p<0.001 to reach significance), we also conducted a Monte Carlo simulation (10,000 runs) to estimate a 95% confidence interval for the number of significant results that could be expected by chance, given the total number of comparisons [35].

4 Results

4.1 Behaviors during Each Affective Episode

Table 2 shows the average frequency of each behavior per affective clip (20 seconds) during episodes of confusion, frustration, engaged concentration, and boredom. A Friedman chi-square analysis reveals statistically significant differences in the frequency of all behaviors across the four affective states. *Inside No Action*—remaining

Table 2: Average number of occurrences of each behavior per clip for each affective state.

Construct	Confusion	Frustration	Concentration	Boredom	Friedman Statistic	p-val
Inside No Action	**0.674**	0.654	0.509	0.286	68.47	0.000
Outside No Action	0.202	0.263	0.387	**0.621**	76.00	0.000
Worksheet	**0.050**	0.043	0.038	0.029	12.44	0.006
Scan After Exploration	**0.048**	0.014	0.015	0.008	10.91	0.012
Long Conversation	0.027	0.022	0.043	**0.044**	36.50	0.000
Repeated Testing	**0.009**	0.002	0.005	0.001	13.15	0.004
Rushing	0.006	0.008	0.013	**0.016**	8.52	0.036

inside a location without performing any action beyond basic movement during a 20-second interval—is the most common behavior during episodes of confusion (0.674), frustration (0.654), and engaged concentration (0.509), but not during boredom (0.286). Instead, *Outside No Action*—the other inactive behavior, occurring outside any game location—is most frequent during boredom (0.621), at a level comparable to *Inside No Action* in the other three affective states. The frequency of *Outside No Action* during confusion (0.202), frustration (0.263), and concentration (0.387) is significantly lower, roughly equivalent to *Inside No Action* during boredom. Although both are inactive behaviors, the affective patterns suggest we might interpret them differently. *Inside No Action* may indicate reaching an impasse (e.g., uncertainty about next steps), requiring students to pause and/or reflect as they work to manage or resolve confusion and frustration. By contrast, *Outside No Action* may represent a deeper disengagement, in which the student chooses to remain outside rather than engage with locations where scientific actions are possible (i.e., inside the various virtual buildings), explaining its higher prevalence during boredom.

Among the more active behavioral categories, notable differences emerged in the use of the *Worksheet* mechanic—a tool that helps students organize and systematize their findings—which was more common during confusion (0.050), frustration (0.043), and, to a lesser extent, engaged concentration (0.038), compared to boredom (0.029). This pattern suggests that students often turn to the worksheet when they reach an impasse and need to structure their findings to consider potential next steps. A similar trend was observed for *Scan After Exploration*—hypothesis testing after gathering information from readings or NPCs—which occurred more frequently during confusion (0.048) and, to a lesser extent, engaged concentration (0.015) and frustration (0.014), than during boredom (0.008). *Repeated Testing* also appears more often during confusion (0.009) and concentration (0.005) than during boredom (0.001), though it was infrequent during frustration (0.002). Both *Scan After Exploration* and *Repeated Testing* reflect hypothesis-testing behaviors that can help resolve confusion or may simply represent normal activities when students are actively engaged.

In contrast, two other active behaviors were more strongly associated with boredom and engaged concentration—two affective states that are not thought to be similar but that typically share low arousal—than with confusion and frustration. *Long Conversations* with NPCs, where students can ask about symptoms, game mechanics, or microbiological concepts, are least frequent during confusion (0.027) and frustration (0.022). Its association with boredom (0.044) and engaged concentration (0.043) suggests students may engage in these conversations (perhaps going through the entire script) when their arousal is low. Likewise, *Rushing*, a behavior that encompasses faster behaviors in both conversations and in-game reading materials, is also more common during boredom (0.016) and engaged concentration (0.013) than during confusion (0.006) or frustration (0.008). This pattern suggests that when arousal is low, students either rush through material (possibly with the intent to cover a lot of content quickly) or slow down (perhaps to better reflect on the content if engaged, or as a signal of boredom and potential disengagement).

4.2 Behaviors Associated with External Learning and Motivational Measures

The seven behaviors included in this study were correlated to the five external measures of learning and motivation. These correlations were calculated first for behaviors that took place across each student's entire gameplay experience with Crystal Island (Section 4.2.1, 35 comparisons). Next, they were calculated for the same behaviors but only analyzing data that took place during each one of the four affective states (Sections 4.2.2 through 4.2.5; $35 \times 4 = 140$ comparisons). Of the 175 total comparisons, 19 are statistically significant ($p < 0.05$), including five correlations for the entire gameplay, seven during confusion, three each during frustration and engaged concentration, and one during boredom episodes (Figure 2). After applying a Benjamini-Hochberg correction for multiple comparisons at the affect level, none of them remained significant. However, a Monte Carlo analysis with a 95% confidence interval shows that only 4 to 15 significant results are likely due to chance. Therefore, the observation of 19 significant correlations is highly unlikely to be due to random variation alone ($p < 0.001$). Although it is possible that some individual results occurred by chance, the overall pattern of findings, as presented below, is unlikely to be attributable to random noise.

4.2.1 Behavioral Associations across the Entirety of Student's Game Play. Table 3 presents the correlations between behaviors aggregated across the entire gameplay and the motivational measures. *Outside No Action*, previously discussed as a potential indicator of roaming or disengagement, is negatively associated with situational interest ($\tau = -0.152$, $p = 0.014$), self-efficacy ($\tau = -0.137$, $p = 0.028$), and pre-test scores ($\tau = -0.128$, $p = 0.048$), while its association with learning gains ($\tau = -0.183$, $p = 0.069$) is marginally significant. All these associations are monotonic according to the GAMs. Students with lower motivation or prior knowledge in science were more likely to roam around the virtual world without engaging in

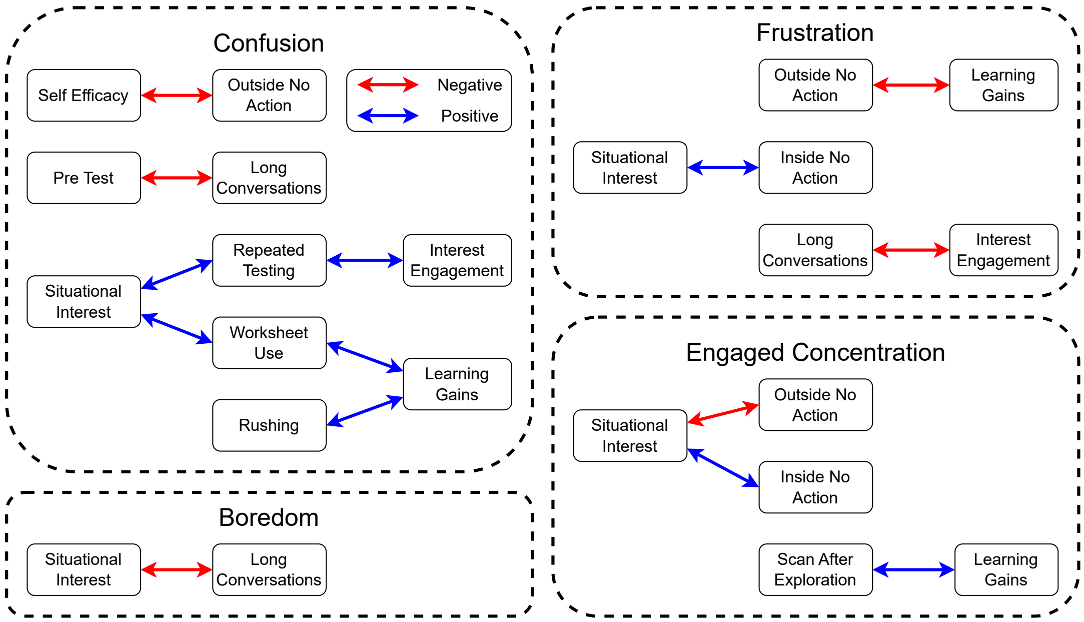

Figure 2: Associations identified between behaviors and motivational and knowledge measures, across affective states.

Table 3: Correlations between behaviors during the entire gameplay and multiple learning and motivational measures. Significant correlations are shown in bold.

	Situational Interest		Self-Efficacy		Interest Engagement		Pre Test		Learning Gains	
Feature	τ	p	τ	p	τ	p	τ	p	τ	p
Outside No Action	**-0.152**	**0.014**	**-0.137**	**0.028**	-0.097	0.250	**-0.128**	**0.048**	-0.183	0.069
Inside No Action	0.109	0.078	0.113	0.070	0.058	0.494	0.101	0.120	0.109	0.281
Rushing	0.101	0.130	0.058	0.387	0.076	0.396	0.030	0.670	0.138	0.196
Long Conversation	-0.049	0.429	-0.054	0.388	-0.075	0.376	-0.096	0.138	0.033	0.742
Worksheet	0.116	0.060	0.079	0.207	0.003	0.970	0.028	0.672	0.151	0.135
Scan After Exploration	0.094	0.173	0.025	0.725	0.052	0.571	0.089	0.219	0.193	0.080
Repeated Testing	**0.151**	**0.031**	0.100	0.157	**0.229**	**0.015**	0.013	0.865	0.197	0.082

science-related actions, which also hinders their potential learning. In contrast, *Repeated Testing*, characterized by rapidly testing multiple hypotheses or submitting concept matrices, was positively associated with both situational interest ($\tau = 0.151$, $p = 0.031$) and interest engagement ($\tau = 0.229$, $p = 0.015$), indicating that students who tested few hypotheses or submitted fewer concept matrices had lower situational interest and interest engagement.

4.2.2 Behavioral Associations that Occur Only during Confusion. We first examine the data for confusion (Table 4), as it has the greatest number of significant correlations (N=7). Students who reported higher situational interest before playing the game are more likely to engage in concrete, purposeful actions during confusion episodes. These included using the *Worksheet* ($\tau = 0.168$, $p = 0.026$) and engaging in *Repeated Testing* ($\tau = 0.173$, $p = 0.031$), both of which show monotonic associations. *Worksheet* use also displays a significant positive association with learning gains ($\tau = 0.269$, $p = 0.019$), underscoring the value of this mechanic for addressing confusion or cognitive impasses by helping students systematize information and reflect. *Rushing* was similarly associated with learning gains ($\tau = 0.245$, $p = 0.044$), suggesting that skimming readings or conversations to locate needed information can also be an effective strategy for overcoming confusion and supporting learning.

Post-test interest-engagement was also positively linked to a concrete action like *Repeated Testing* ($\tau = 0.261$, $p = 0.015$) but in a non-monotonic way according to the GAM (degree 1.83). Students who engaged in moderate levels of *Repeated Testing* reported the highest interest engagement, whereas low and high levels of testing corresponded to lower engagement, with students low in interest also performing the least testing. This pattern suggests that some students, when confused, remained engaged and adopted a trial-and-error approach to solving the game's challenges, which may involve limited reflection in the moment but can help sustain their engagement. However, excessive testing may also signal declining engagement and a shift toward less constructive behaviors.

Table 4: Correlations between actions during confusion episodes and multiple learning and motivational measures. Significant correlations are shown in bold.

	Situational Interest		Self-Efficacy		Interest Engagement		Pre Test		Learning Gains	
Feature	τ	p	τ	p	τ	p	τ	p	τ	p
Outside No Action	-0.109	0.134	**-0.173**	**0.019**	0.105	0.293	-0.057	0.460	0.044	0.701
Inside No Action	0.026	0.714	0.050	0.486	-0.161	0.091	0.071	0.342	-0.178	0.104
Rushing	0.141	0.077	0.035	0.669	0.108	0.310	-0.027	0.749	**0.245**	**0.044**
Long Conversation	0.101	0.183	-0.025	0.745	0.166	0.103	**-0.171**	**0.033**	0.146	0.206
Worksheet	**0.168**	**0.026**	0.123	0.110	0.045	0.658	-0.086	0.281	**0.269**	**0.019**
Scan After Exploration	0.122	0.120	0.053	0.509	0.137	0.193	0.045	0.590	0.099	0.414
Repeated Testing	**0.173**	**0.031**	0.158	0.053	**0.261**	**0.015**	-0.031	0.713	0.166	0.174

Likewise, high self-efficacy was linked with more concrete actions during confusion. Those students with lower self-efficacy, in contrast, spent extended periods of time roaming outside (*Outside No Action*) without participating in science-related activities ($\tau = -0.173$, $p = 0.019$), indicating that students with the lowest confidence in their scientific abilities may struggle to identify or initiate effective strategies to resolve confusion. Similarly, there is a negative monotonic association between pre-test scores and *Long Conversations* with NPCs ($\tau = -0.171$, $p = 0.033$), suggesting that students with less prior knowledge may find it harder to skim conversations and extract the specific information needed to address their confusion.

Confusion stands out with the highest number of significant correlations (N=7), including three (out of five) that align with results from the aggregated gameplay data (see Table 3, above). This suggests that the behaviors during moments of confusion may drive much of the overall correlation observed in the full gameplay data. More broadly, confusion may be the state in which motivational factors and prior knowledge exert the strongest influence on students' behaviors, and in which behaviors, in turn, might have a strong impact on learning outcomes.

4.2.3 Behavioral Associations that Occur Only during Frustration. Table 5 shows the correlations between student actions during frustration episodes and their learning and motivational measures. All three of the significant associations for frustration episodes were monotonic. For students with high learning gains, *Outside No Action* is uncommon during frustration ($\tau = -0.256$, $p = 0.042$). Analysis also shows that it is uncommon for students with high situational interest, though this relationship is only marginally significant ($\tau = -0.151$, $p = 0.057$). This behavior is required to get between sites of the game where more concrete scientific actions can be conducted, but it is impossible to engage with scientific material while outside in Crystal Island. Therefore, it is unsurprising that spending considerable amounts of time outside does not lead to learning gains, though it is notable that the relationship between *Outside No Action* and negative learning gains was only marginally significant when considering all of the data (Table 3). Students with high interest engagement are also unlikely to participate in *Long Conversations* during frustration ($\tau = -0.229$, $p = 0.043$).

In contrast, for students with higher situational interest, *Inside No Action* is more common during frustration ($\tau = 0.176$, $p = 0.020$). In general, the pauses in action captured by *Inside No Action* could indicate disengagement. However, the marginally positive association between *Inside No Action* and learning gains ($\tau = 0.225$, $p = 0.058$), suggests that this lack of action (unlike roaming outside) can be indicative of self-regulation strategies that are eventually beneficial for learning.

4.2.4 Behavioral Associations that Occur Only during Engaged Concentration. Table 6 shows the correlations between student actions during engaged concentration episodes and their learning and motivational measures, three of which were statistically significant. All associations were monotonic according to the GAMs, and two of the behaviors that showed statistically significant reflect findings from either the full data set or the analysis of another affective state. High situational interest is correlated with low levels of *Outside No Action* during engaged concentration ($\tau = -0.256$, $p = 0.042$), just as it was when the entire data set was analyzed. Likewise, this association is also marginally significant for prior knowledge and for learning gains ($\tau = -0.126$, $p = 0.056$) and lower learning gains ($\tau = -0.191$, $p = 0.062$), suggesting difficulties stemming from insufficient knowledge to engage more effectively with the game. These results also somewhat mirror the analysis of the full data set, where *Outside No Action* is negatively associated with both prior knowledge and self-efficacy. Situational interest is also correlated with high levels of *Inside No Action* during engaged concentration ($\tau = 0.184$, $p = 0.004$), just as it was for frustration, probably indicating again moments when students take a pause to reflect on their own process or next steps to solve the mystery of the game.

Learning gains are positively associated with *Scan After Exploration* during concentration ($\tau = -0.242$, $p = 0.038$). As this behavior reflects direct engagement with the scientific content of the game, this is an expected and desired result. However, it is notable that this behavior is only positively associated with learning gains when it occurs during concentration episodes, without any other significant correlation with outcomes, even when the full data set is analyzed. This finding underscores the importance of maintaining focus for this exploration-scanning cycle.

4.2.5 Behavioral Associations that Occur Only during Boredom. Table 7 presents the correlations between student actions during boredom episodes and their learning and motivational measures. Situational interest is associated with fewer *Long Conversations* during boredom ($\tau = -0.126$, $p = 0.047$), but the GAM indicates that this negative association follows a decreasing inverted U-shape (degree=1.82). This non-monotonic finding shows that situational

Table 5: Correlations between actions during frustration episodes and multiple learning and motivational measures. Significant correlations are shown in bold.

	Situational Interest		Self-Efficacy		Interest Engagement		Pre Test		Learning Gains	
Feature	τ	p	τ	p	τ	p	τ	p	τ	p
Outside No Action	-0.151	0.057	-0.047	0.561	0.027	0.804	-0.026	0.754	**-0.256**	**0.042**
Inside No Action	**0.176**	**0.020**	0.136	0.077	0.172	0.094	-0.003	0.965	0.225	0.058
Rushing	0.082	0.338	-0.019	0.828	0.066	0.566	-0.167	0.063	-0.032	0.812
Long Conversation	-0.149	0.070	-0.139	0.097	**-0.229**	**0.043**	0.021	0.811	-0.155	0.235
Worksheet	-0.013	0.872	-0.091	0.274	-0.117	0.291	-0.036	0.677	-0.069	0.592
Scan After Exploration	-0.020	0.817	-0.050	0.566	-0.154	0.182	0.157	0.084	-0.013	0.922
Repeated Testing	-0.027	0.755	-0.100	0.253	-0.067	0.567	-0.124	0.172	0.004	0.974

Table 6: Correlations between actions during concentration episodes and multiple learning and motivational measures. Significant correlations are shown in bold.

	Situational Interest		Self-Efficacy		Interest Engagement		Pre Test		Learning Gains	
Feature	τ	p	τ	p	τ	p	τ	p	τ	p
Outside No Action	**-0.173**	**0.006**	-0.107	0.090	-0.130	0.129	-0.126	0.056	-0.191	0.062
Inside No Action	**0.184**	**0.004**	0.110	0.086	0.118	0.173	0.110	0.098	0.125	0.229
Rushing	0.060	0.396	0.075	0.296	0.009	0.925	0.081	0.278	0.223	0.054
Long Conversation	0.037	0.574	0.011	0.868	0.091	0.301	0.002	0.981	0.141	0.185
Worksheet	0.037	0.582	0.080	0.245	-0.076	0.408	0.074	0.302	0.110	0.321
Scan After Exploration	0.052	0.471	0.038	0.605	0.134	0.168	0.040	0.598	**0.242**	**0.038**
Repeated Testing	0.039	0.589	0.020	0.791	0.037	0.711	-0.011	0.887	0.189	0.112

Table 7: Correlations between actions during boredom episodes and multiple learning and motivational measures. Significant correlations are shown in bold.

	Situational Interest		Self-Efficacy		Interest Engagement		Pre Test		Learning Gains	
Feature	τ	p	τ	p	τ	p	τ	p	τ	p
Outside No Action	-0.032	0.607	-0.093	0.138	0.012	0.890	-0.082	0.211	-0.142	0.165
Inside No Action	0.003	0.962	0.062	0.340	0.016	0.855	0.050	0.468	0.073	0.500
Rushing	0.111	0.122	0.064	0.380	0.157	0.109	0.101	0.181	-0.128	0.277
Long Conversation	**-0.126**	**0.047**	-0.009	0.891	-0.143	0.095	0.023	0.734	-0.004	0.971
Worksheet	0.040	0.536	-0.001	0.984	0.036	0.684	0.118	0.087	-0.030	0.786
Scan After Exploration	0.076	0.301	0.055	0.466	0.027	0.787	-0.004	0.958	0.140	0.249
Repeated Testing	0.078	0.296	0.035	0.642	0.088	0.386	-0.007	0.932	0.100	0.417

interest is highest among students who engaged in moderate levels of long conversations reported the highest situational interest, whereas those with lower situational interest either did not engage NPCs at all or had extremely *Long Conversations*. No other significant associations were identified during boredom episodes.

5 Discussion and Conclusion

In this study, we investigated the relationship between the behaviors adopted by students during different affective states and external measures of knowledge and motivation, identifying how the choices students make in response to their affect shape their outcomes. While some behaviors occurred frequently throughout the entire gameplay, their prevalence shifted depending on the emotion students were experiencing. For instance, during episodes of boredom, students often roamed outside locations without performing science-related actions—a behavior that, when aggregated across gameplay, was negatively correlated with both learning and motivational measures. In contrast, when not bored, students were more likely to remain inside locations and engage in actions potentially linked to self-regulated learning [25], such as systematizing information in the worksheet, testing hypotheses after having explored, or simply pausing to reflect. In fact, even pauses that occurred inside the virtual locations of the game were positively associated with some motivation and learning under certain affective conditions.

Previous research has shown relatively uniform associations between boredom and learning outcomes in a variety of systems [28]. In the context of this game, boredom has already been linked to lower levels of self-efficacy, situational interest, and learning gains [55]. The stronger prevalence of *Outside no Action* during boredom and the negative associations between this behavior and multiple motivational and learning measures suggests a pathway that explains these relationships, where students with low interest or self-efficacy are more likely to experience boredom, which results in roaming behaviors that ultimately undermine learning. Additionally, the lack of multiple significant associations between behaviors and outcomes for boredom suggests that student behavior in this state is relatively uniform and less variable. Thus, beyond other specific actions students could take, the very experience of boredom, and the roaming it prompts, appears to undermine students' ability to achieve the intended outcomes of the game. This interpretation aligns with the predictions of the CVT and SDVET models [38, 43], extending them by identifying the off-task behavior associated with boredom as a key driver of reduced learning outcomes.

In contrast, when students experience confusion, frustration, or engaged concentration, they tend to remain involved in scientific actions beyond simple roaming, or at least stay within locations where such actions can be performed. The greater behavioral variability observed in these states (particularly in confusion) may reflect a more complex typology (or constellation of types) for these emotions, in line with past models that have made distinctions between canonical and pleasant frustration [22], intolerable and tolerable confusion (SDVET; [38]), and the *confrustion* constellation [2]. Our findings align with the SDVET model, which suggests that students' tolerance for impasses may be influenced by external motivation factors like self-efficacy and interest, which in turn mediate both their behavioral responses to these emotions and the ultimate learning outcomes they achieve.

Within episodes of confusion, self-efficacy plays an important role in preventing students from roaming and experience intolerable confusion. Students with lower perceived capacity for conducting scientific tasks were more likely to disengage and move outside locations, whereas students with higher self-efficacy remained inside, having more opportunities to perform actions that could help them resolve the cognitive disequilibrium that triggered their confusion and, in turn, foster learning. Sabourin and Lester [50] argued that staying on-task during confusion supports learning. Our findings extend this by showing that, although all five science-related actions had positive associations with learning gains when performed during confusion, the two most (significantly) beneficial behaviors were (a) skimming readings or NPC interactions to quickly locate the information needed to address confusion, and (b) using the worksheet as a reflection tool. Both represent key self-regulatory strategies and were primarily motivated by higher situational interest. In this way, we build on Sabourin and Lester's findings by showing that remaining on task during confusion is most beneficial when students recognize which concrete actions can help resolve their confusion and are motivated to pursue them. By contrast, less efficient behaviors, such as engaging in overly long conversations with NPCs, were more common among students with lower prior knowledge, suggesting that these students may not recognize which specific actions are effective for addressing confusion, or that they may be experiencing deeper cognitive struggle.

Frustration episodes also reveal wide variability in learning and motivational measures depending on how students behave. Consistent with Sabourin and Lester's [50] previous research (and with broader evidence from non-game educational software [5]), our findings indicate that off-task behavior is not always detrimental when it emerges in response to unpleasant emotions. Although Sabourin and Lester categorized any non-scientific action as off-task, in this study, we distinguish between remaining within task-relevant locations and wandering outside them. This finer categorization allowed us to observe that when off-task behavior involved staying in a location without acting, influenced by a high situational interest, students may achieve higher learning outcomes. In these cases, inactivity may actually reflect intentional pauses for self-regulation, such as reflecting on prior findings or planning next steps. By contrast, when situational interest is absent, students are more likely to disengage by leaving task-relevant locations, suggesting a deeper cognitive struggle or broader difficulties with self-regulation and motivation,n and aligning more closely with canonical unpleasant frustration and poorer learning outcomes.

For episodes of engaged concentration, situational interest also plays a key role in shaping students' actions. As observed for frustration, students with higher situational interest are more likely to remain inside task-relevant locations, whereas those with lower interest or prior knowledge tend to roam outside. Staying within locations during concentration—particularly testing hypotheses, one of the core scientific actions of the game—is associated with higher learning outcomes. In contrast, leaving these locations and disengaging from scientific actions appears to hinder learning, matching the pattern for frustration. This suggests that simply being engaged with the game is not enough. Without interest or knowledge to guide constructive behaviors, engagement may not translate into meaningful learning.

The differences in correlations observed for confusion and frustration, along with the similarities between frustration and engaged concentration, carry important implications for how these emotions are analyzed within theoretical frameworks and future research. For instance, Baker et al. [2] have argued that confusion and frustration should not be treated as independent affective states but rather as a constellation of multiple subtypes, where some forms of frustration may resemble certain forms of confusion more closely than alternative forms of frustration. Our results align with this model, suggesting that both confusion and frustration can arise from multiple causes and have multiple manifestations. For example, low self-efficacy or situational interest may lead to unpleasant or unconstructive forms of frustration or confusion that hinder learning, as also described by the SDVET model. Conversely, when confusion or frustration is accompanied by high interest, self-efficacy, sufficient knowledge, or effective self-regulatory skills, these states may shift toward a more pleasant or constructive form, resembling more pleasurable emotions such as engaged concentration. At the same time, our findings underscore that confusion and frustration cannot be treated as interchangeable. Therefore, studies that conceptualize *confrustion* as a single emotion or implement machine-learned *confrustion* detectors should proceed with caution, recognizing that multiple subtypes of these emotions with distinct manifestations are being grouped together.

No study is without limitations. Factors such as the topic, game mechanics, or age group may influence these results. Therefore, although our findings align with many theoretical predictions from prior models and match several results reported in the literature, they reflect only one specific context and should be replicated across multiple settings. Additionally, some of the behavioral categories used in this study may have conflated distinct phenomena. For instance, remaining inside a location without performing any action could signal reflection but could also indicate disengagement. Future research should incorporate qualitative triangulation, employing data collection methods such as data-driven classroom interviews [3, 39], that allow researchers to better understand why students engaged in particular actions or behaviors. Furthermore, both detectors and behavioral codes are derived from interaction logs, which introduces a potential risk of circularity. However, the fact that the detectors and codes operate at different levels of abstraction, along with the correlations observed with external variables, reduces this concern and supports the validity of our findings. Even so, additional qualitative triangulation could further mitigate this issue. Lastly, although the detectors used in this study were tested for algorithmic bias, our ability to determine how well their performance generalizes is limited for less represented groups or intersections of groups (e.g., Latino boys or Asian girls) due to the low number of students (less than ten) in these groups [58]. Moreover, even when algorithms show no evidence of biased performance, cultural differences in how emotions are expressed and understood may affect the ground truth itself. This is particularly relevant for students whose cultural background differs from that of the observers [14, 37] or for those from cultures in which expressing certain emotions is socially discouraged [26, 31, 42].

In conclusion, these findings contribute to a more nuanced understanding of how affective states shape learning through the behaviors they elicit. Whereas boredom appears relatively stable in its (negative) associations with outcomes, engaged concentration, frustration, and mainly confusion demonstrate greater variability, with their impact depending on students' initial motivation, self-efficacy, and the strategies they adopt in response. This study highlights the importance of examining not only which emotions students experience but also how they translate those emotions into concrete actions. These insights extend existing frameworks by providing evidence on the behavioral mechanisms through which affect operates, suggesting ways to nudge student behavior so that they can transform moments of confusion and frustration into constructive pathways for engagement and growth.

Acknowledgments

This research was supported by National Science Foundation Grants IIS-2016943 and IIS-2016993. Andres Felipe Zambrano thanks the Ministerio de Ciencia, Tecnología e Innovación and the Fulbright-Colombia commission for supporting his doctoral studies through the Fulbright-MinCiencias 2022 scholarship.

References

[1] Amaël Arguel, Lori Lockyer, Ottmar V. Lipp, Jason M. Lodge, and Gregor Kennedy. 2017. Inside Out: Detecting Learners' Confusion to Improve Interactive Digital Learning Environments. *Journal of Educational Computing Research* 55, 4 (July 2017), 526–551. doi:10.1177/0735633116674732 Publisher: SAGE Publications Inc.

[2] Ryan S. Baker, Elizabeth Cloude, Juliana Andres, and Zhanlan Wei. 2024. The Confrustion Constellation: A New Way of Looking at Confusion and Frustration. (June 2024). https://osf.io/preprints/edarxiv/4rgtk

[3] Ryan S. Baker, Stephen Hutt, Nigel Bosch, Jaclyn Ocumpaugh, Gautam Biswas, Luc Paquette, J. M. Alexandra Andres, Nidhi Nasiar, and Anabil Munshi. 2024. Detector-driven classroom interviewing: focusing qualitative researche r time by selecting cases in situ. *Educational technology research and development* 72, 5 (Oct. 2024), 2841–2863. doi:10.1007/s11423-023-10324-y

[4] Ryan S. J. d. Baker, Sidney K. D'Mello, Ma.Mercedes T. Rodrigo, and Arthur C. Graesser. 2010. 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. *International Journal of Human-Computer Studies* 68, 4 (April 2010), 223–241. doi:10.1016/j.ijhcs.2009.12.003

[5] Ryan S. J. d. Baker, Gregory R. Moore, Angela Z. Wagner, Jessica Kalka, Aatish Salvi, Michael Karabinos, Colin A. Ashe, and David Yaron. 2011. The Dynamics between Student Affect and Behavior Occurring Outside of Educational Software. In *Affective Computing and Intelligent Interaction*, Sidney D'Mello, Arthur Graesser, Björn Schuller, and Jean-Claude Martin (Eds.). Springer Berlin Heidelberg, Berlin, Heidelberg, 14–24.

[6] Ryan S. J. d. Baker and Jaclyn Ocumpaugh. 2014. Interaction-Based Affect Detection in Educational Software. *The Oxford handbook of affective computing* (2014), 233. Publisher: Oxford University Press.

[7] Ryan S. J. d. Baker, Ma. Mercedes T. Rodrigo, and Ulises E. Xolocotzin. 2007. The Dynamics of Affective Transitions in Simulation Problem-Solving Environments. In *Affective Computing and Intelligent Interaction*, Ana C. R. Paiva, Rui Prada, and Rosalind W. Picard (Eds.). Springer Berlin Heidelberg, Berlin, Heidelberg, 666–677.

[8] Yoav Benjamini and Yosef Hochberg. 1995. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society: Series B (Methodological)* 57, 1 (1995), 289–300. doi:10.1111/j.2517-6161.1995.tb02031.x _eprint: https://rss.onlinelibrary.wiley.com/doi/pdf/10.1111/j.2517-6161.1995.tb02031.x.

[9] Nigel Bosch and Sidney D'Mello. 2017. The Affective Experience of Novice Computer Programmers. *International Journal of Artificial Intelligence in Education* 27, 1 (March 2017), 181–206. doi:10.1007/s40593-015-0069-5

[10] Anthony F Botelho, Ryan S Baker, Jaclyn Ocumpaugh, and Neil T Heffernan. 2018. Studying Affect Dynamics and Chronometry Using Sensor-Free Detectors.. In *International Conference on Educational Data Mining*. ERIC, 157–166.

[11] Shari L. Britner and Frank Pajares. 2006. Sources of science self-efficacy beliefs of middle school students. *Journal of Research in Science Teaching* 43, 5 (2006), 485–499. doi:10.1002/tea.20131 _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/tea.20131.

[12] Cyril Brom, Filip Děchtěrenko, Nikola Frollová, Tereza Stárková, Edita Bromová, and Sidney K. D'Mello. 2017. Enjoyment or involvement? Affective-motivational mediation during learning from a complex computerized simulation. *Computers & Education* 114 (Nov. 2017), 236–254. doi:10.1016/j.compedu.2017.07.001

[13] Elizabeth B. Cloude, Daryn A. Dever, Debbie L. Hahs-Vaughn, Andrew J. Emerson, Roger Azevedo, and James Lester. 2022. Affective Dynamics and Cognition During Game-Based Learning. *IEEE Transactions on Affective Computing* 13, 4 (2022), 1705–1717. doi:10.1109/TAFFC.2022.3210755

[14] Cristina Conati and Mirela Gutica. 2016. Interaction with an Edu-Game: A Detailed Analysis of Student Emotions and Judges' Perceptions. *International Journal of Artificial Intelligence in Education* 26, 4 (Dec. 2016), 975–1010. doi:10.1007/s40593-015-0081-9

[15] Mihaly Csikszentmihalyi. 1990. *Flow: The psychology of optimal experience*. Vol. 1990. Harper & Row New York.

[16] Edward L. Deci and Richard M. Ryan. 2007. SDT: Questionnaires: Intrinsic motivation inventory (IMI). *Retrieved October* 27 (2007), 2009.

[17] Sidney D'Mello and Caitlin Mills. 2014. Emotions while writing about emotional and non-emotional topics. *Motivation and Emotion* 38, 1 (Feb. 2014), 140–156. doi:10.1007/s11031-013-9358-1 Place: New York Publisher: Springer Nature B.V..

[18] Sidney K. D'Mello, Blair Lehman, and Natalie Person. 2010. Monitoring Affect States During Effortful Problem Solving Activities. *International Journal of Artificial Intelligence in Education* 20, 4 (Nov. 2010), 361–389. doi:10.3233/JAI-2010-012 Publisher: SAGE Publications.

[19] Sidney D'Mello. 2017. Emotional Learning Analytics. In *The Handbook of Learning Analytics* (1 ed.), Charles Lang, George Siemens, Alyssa Friend Wise, and Dragan Gaševic (Eds.). Society for Learning Analytics Research (SoLAR), Alberta, Canada, 115–127. http://solaresearch.org/hla-17/hla17-chapter1

[20] Sidney D'Mello and Art Graesser. 2012. Dynamics of affective states during complex learning. *Learning and Instruction* 22, 2 (2012), 145–157. doi:10.1016/j.learninstruc.2011.10.001

[21] Stephen Fancsali. 2014. Causal discovery with models: behavior, affect, and learning in cognitive tutor algebra. In *Educational Data Mining 2014*. Citeseer.

[22] James Paul Gee. 2005. Learning by Design: Good Video Games as Learning Machines. *E-Learning and Digital Media* 2, 1 (March 2005), 5–16. doi:10.2304/elea.2005.2.1.5 Publisher: SAGE Publications.

[23] Lilin Gong, Yang Liu, and Wei Zhao. 2019. Dynamics of Emotional States and Their Relationship with Learning Outcomes during Learning Python with MOOC. In *Proceedings of the 2019 7th International Conference on Information and Education Technology (ICIET 2019)*. Association for Computing Machinery, New York, NY, USA, 71–76. doi:10.1145/3323771.3323821 event-place: Aizu-Wakamatsu, Japan.

[24] Arthur Graesser, Sidney D'Mello, Patrick Chipman, Brandon King, and Bethany McDaniel. 2007. Exploring relationships between affect and learning with AutoTutor. In *Proceedings of International Conference of Artificial Intelligence in Education*.

[25] Jeffrey A Greene, Matthew L Bernacki, and Allyson F Hadwin. 2023. Self-regulation. *Handbook of educational psychology* (2023), 314–334. Publisher: Routledge.

[26] Jason Matthew Harley. 2016. Measuring emotions: A survey of cutting edge methodologies used in com puter-based learning environment research. *Emotions, technology, design, and learning* (2016), 89–114.

[27] Jon-Chao Hong, Ming-Yueh Hwang, Kai-Hsin Tai, and Pei-Hsin Lin. 2021. The effects of intrinsic cognitive load and gameplay interest on flow experience reflecting performance progress in a Chinese remote association game. *Computer Assisted Language Learning* 34, 3 (May 2021), 358–378. doi:10.1080/09588221.2019.1614068 Publisher: Routledge.

[28] Shamya Karumbaiah, Ryan Baker, Yan Tao, and Ziyang Liu. 2022. How does Students' Affect in Virtual Learning Relate to Their Outcomes ? A Systematic Review Challenging the Positive-Negative Dichotomy. In *LAK22: 12th International Learning Analytics and Knowledge Conference (LAK22)*. Association for Computing Machinery, New York, NY, USA, 24–33. doi:10.1145/3506860.3506863

[29] Maurice G Kendall. 1938. A new measure of rank correlation. *Biometrika* 30, 1-2 (1938), 81–93. Publisher: Oxford University Press.

[30] Victor Kostyuk, Ma. Victoria Almeda, and Ryan S. Baker. 2018. Correlating affect and behavior in reasoning mind with state test achievement. In *Proceedings of the 8th International Conference on Learning Analytics and Knowledge (LAK '18)*. Association for Computing Machinery, New York, NY, USA, 26–30. doi:10.1145/3170358.3170378 event-place: Sydney, New South Wales, Australia.

[31] Jon A. Krosnick. 1999. Survey Research. *Annual Review of Psychology* 50, Volume 50, 1999 (1999), 537–567. doi:10.1146/annurev.psych.50.1.537

[32] Diane Marie C. Lee, Ma. Mercedes T. Rodrigo, Ryan S. J. d. Baker, Jessica O. Sugay, and Andrei Coronel. 2011. Exploring the Relationship between Novice Programmer Confusion and Achievement. In *Affective Computing and Intelligent Interaction*, Sidney D'Mello, Arthur Graesser, Björn Schuller, and Jean-Claude Martin (Eds.). Springer Berlin Heidelberg, Berlin, Heidelberg, 175–184.

[33] Lisa Linnenbrink-Garcia, Amanda M Durik, Anne Marie M Conley, Kenneth E Barron, John M Tauer, Stuart A Karabenick, and Judith M Harackiewicz. 2010. Measuring situational interest in academic domains. *Educational and psychological measurement* 70, 4 (2010), 647–671. doi:10.1177/0013164409355699 Publisher: Sage Publications Sage CA: Los Angeles, CA.

[34] Zhongxiu Liu, Visit Pataranutaporn, Jaclyn Ocumpaugh, and Ryan Baker. 2013. Sequences of frustration and confusion, and learning. In *International Conference on Educational Data Mining*. 114–120.

[35] Nicholas Metropolis and S. Ulam. 1949. The Monte Carlo Method. *J. Amer. Statist. Assoc.* 44, 247 (Sept. 1949), 335–341. doi:10.1080/01621459.1949.10483310 Publisher: ASA Website.

[36] Nidhi Nasiar, Andres Felipe Zambrano, Jaclyn Ocumpaugh, Alex Goslen, Jonathan Rowe, Jessica Vandenberg, Jordan Esiason, and Stephen Hutt. 2024. The Influence of Different Measurement Approaches on Student Affect Tr ansitions Using Ordered Networks. In *Advances in Quantitative Ethnography*, Yoon Jeon Kim and Zachari Swiecki (Eds.). Springer Nature Switzerland, Cham, 195–203.

[37] Jaclyn Ocumpaugh, Ryan S. Baker, and Ma Mercedes T Rodrigo. 2015. Baker Rodrigo Ocumpaugh Monitoring Protocol (BROMP) 2.0 Technical and Training Manual. (2015).

[38] Jaclyn Ocumpaugh, Nidhi Nasiar, Andres Felipe Zambrano, Alex Goslen, Jessica Vandenberg, Jordan Esiason, Jonathan Rowe, and Stepehn Hutt. 2025. Refocusing the lens through which we view affect dynamics: The Skills, Difficulty, Value, Efficacy and Time Model. Dublin, Ireland.

[39] Jaclyn Ocumpaugh, Luc Paquette, Ryan S. Baker, Amanda Barany, Jeff Ginger, Nathan Casano, Andres F. Zambrano, Xiner Liu, Zhanlan Wei, Yiqui Zhou, Qianhui Liu, Stephen Hutt, Alexandra M. A. Andres, Nidhi Nasiar, Camille Giordano, Martin van Velsen, and Micheal Mogessi. 2025. The Quick Red Fox gets the best Data Driven Classroom Interviews: A manual for an interview app and its associated methodology. arXiv:2511.13466 [cs.HC] https://arxiv.org/abs/2511.13466

[40] Gustavo Padron-Rivera, Cristina Joaquin-Salas, Jose-Luis Patoni-Nieves, and Juan-Carlos Bravo-Perez. 2018. Patterns in Poor Learning Engagement in Students While They Are Solving Mathematics Exercises in an Affective Tutoring System Related to Frustration. In *Pattern Recognition*, José Francisco Martínez-Trinidad, Jesús Ariel Carrasco-Ochoa, José Arturo Olvera-López, and Sudeep Sarkar (Eds.). Springer International Publishing, Cham, 169–177.

[41] Zachary A Pardos, Ryan SJD Baker, Maria OCZ San Pedro, Sujith M Gowda, and Supreeth M Gowda. 2014. Affective States and State Tests: Investigating How Affect and Engagement during the School Year Predict End-of-Year Learning Outcomes. *Journal of Learning Analytics* 1, 1 (2014), 107–128. Publisher: ERIC.

[42] Delroy L. Paulhus. 1991. Measurement and control of response bias. In *Measures of personality and social psychological attitudes*. Academic Press, San Diego, CA, US, 17–59.

[43] Reinhard Pekrun. 2006. The Control-Value Theory of Achievement Emotions: Assumptions, Corollaries, and Implications for Educational Research and Practice. *Educational Psychology Review* 18, 4 (Dec. 2006), 315–341. doi:10.1007/s10648-006-9029-9

[44] Stephen R. Porter, Michael E. Whitcomb, and William H. Weitzer. 2004. Multiple surveys of students and survey fatigue. *New Directions for Institutional Research* 2004, 121 (Jan. 2004), 63–73. doi:10.1002/ir.101 Publisher: John Wiley & Sons, Ltd.

[45] J. Elizabeth Richey, Juan Miguel L. Andres-Bray, Michael Mogessie, Richard Scruggs, Juliana M. A. L. Andres, Jon R. Star, Ryan S. Baker, and Bruce M. McLaren. 2019. More confusion and frustration, better learning: The impact of erroneous examples. *Computers & Education* 139 (2019), 173–190. doi:10.1016/j.compedu.2019.05.012

[46] J. Elizabeth Richey, Bruce M. McLaren, Miguel Andres-Bray, Michael Mogessie, Richard Scruggs, Ryan Baker, and Jon Star. 2019. Confrustion in Learning from Erroneous Examples: Does Type of Prompted Self-explanation Make a Difference?. In *Artificial Intelligence in Education*, Seiji Isotani, Eva Millán, Amy Ogan, Peter Hastings, Bruce McLaren, and Rose Luckin (Eds.). Springer International Publishing, Cham, 445–457.

[47] Ma. Mercedes T. Rodrigo, Ryan S. Baker, Matthew C. Jadud, Anna Christine M. Amarra, Thomas Dy, Maria Beatriz V. Espejo-Lahoz, Sheryl Ann L. Lim, Sheila A.M.S. Pascua, Jessica O. Sugay, and Emily S. Tabanao. 2009. Affective and behavioral predictors of novice programmer achievement. In *Proceedings of the 14th Annual ACM SIGCSE Conference on Innovation and Technology in Computer Science Education (ITiCSE '09)*. Association for Computing Machinery, New York, NY, USA, 156–160. doi:10.1145/1562877.1562929 event-place: Paris, France.

[48] Ma Mercedes T Rodrigo and Ryan S. J. Baker. 2011. Comparing learners' affect while using an intelligent tutor and an edu cational game. *Research and Practice in Technology Enhanced Learning* 6, 1 (March 2011), 43–66.

[49] Jonathan P. Rowe, Lucy R. Shores, Bradford W. Mott, and James C. Lester. 2011. Integrating Learning, Problem Solving, and Engagement in Narrative-Centered Learning Environments. *International Journal of Artificial Intelligence in Education* 21, 1-2 (2011), 115–133. doi:10.3233/JAI-2011-019 Publisher: IOS Press.

[50] Jennifer L Sabourin and James C Lester. 2013. Affect and engagement in Game-Based Learning environments. *IEEE transactions on affective computing* 5, 1 (2013), 45–56.

[51] Maria Ofelia San Pedro, Ryan Baker, Alex Bowers, and Neil Heffernan. 2013. Predicting college enrollment from student interaction with an intelligent tutoring system in middle school. In *International Conference on Educational Data Mining*.

[52] Robert Sawyer, Jonathan Rowe, Roger Azevedo, and James Lester. 2018. Filtered Time Series Analyses of Student Problem-Solving Behaviors in Game-Based Learning. *International Educational Data Mining Society* (2018). Publisher: ERIC.

[53] Valerie J. Shute, Sidney D'Mello, Ryan Baker, Kyunghwa Cho, Nigel Bosch, Jaclyn Ocumpaugh, Matthew Ventura, and Victoria Almeda. 2015. Modeling how incoming knowledge, persistence, affective states, and in -game progress influence student learning from an educational game. *Computers & Education* 86 (Aug. 2015), 224–235. doi:10.1016/j.compedu.2015.08.001

[54] Andres Felipe Zambrano, Amanda Barany, Jaclyn Ocumpaugh, Nidhi Nasiar, Jessica Vandenberg, Alex Goslen, Jordan Esiason, Jonathan Rowe, and Stephen Hutt. 2025. Unlocking Gameplay Insights with Epistemic (Ordered) Network Analysis: Understanding the Potential of Video Games to Foster Authentic Scientific Practices in STEM Education. *Journal of Science Education and Technology* 31 (March 2025), 1164–1177. doi:10.1007/s10956-025-10213-4

[55] Andres Felipe Zambrano, Nidhi Nasiar, Jaclyn Ocumpaugh, Alex Goslen, Jiayi Zhang, Jonathan Rowe, Jordan Esiason, Jessica Vandenberg, and Stephen Hutt. 2024. Says Who? How different ground truth measures of emotion impact student affective modeling. In *Proceedings of the 17th International Conference on Educational Data Mining*, Benjamin Paaßen and Carrie Demmans Epp (Eds.). International Educational Data Mining Society, Atlanta, Georgia, USA, 211–223. doi:10.5281/zenodo.12729800

[56] Andres Felipe Zambrano, Jaclyn Ocumpaugh, Ryan S. Baker, Kirk Vanacore, Jordan Esiason, and Jessica Vandenberg. 2025. The Half-Life of Epistemic Emotions: How Motivation Influences Affective Chronometry. In *Proceedings of the 18th International Conference on Educational Data Mining*. International Educational Data Mining Society, Palermo, Italy.

[57] Andres Felipe Zambrano, Jaclyn Ocumpaugh, Ryan S. Baker, and Jessica Vandenberg. 2026. Better to Be Confused or Frustrated Than Bored: Analyzing Affect Dynamics Across Player Archetypes. In *Advances in Quantitative Ethnography*, Guadalupe Carmona, Cynthia Lima, María Josefa Santos, Héctor Benítez, Luis Montero-Moguel, and Beatriz Galarza-Tohen (Eds.). Springer Nature Switzerland, Cham, 384–399.

[58] Andres Felipe Zambrano, Jiayi Zhang, and Ryan S. Baker. 2024. Investigating Algorithmic Bias on Bayesian Knowledge Tracing and Carelessness Detectors. In *Proceedings of the 14th Learning Analytics and Knowledge Conference (LAK '24)*. Association for Computing Machinery, New York, NY, USA, 349–359. doi:10.1145/3636555.3636890 event-place: Kyoto, Japan.