Better to be Confused or Frustrated than Bored: Analyzing Affect Dynamics Across Player Archetypes

Andres Felipe Zambrano[1[0000-0003-0692-1209]], Jaclyn Ocumpaugh[1[0000-0002-9667-8523]], Ryan S. Baker[1[0000-0002-3051-3232]], Jessica Vandenberg[2[0000-0001-6497-1840]]

[1] University of Pennsylvania, Philadelphia, PA 19104, USA
[2] North Carolina State University, Raleigh, NC 27695, USA
azamb13@upenn.edu

Abstract. Ordered Network Analysis (ONA) helps us to understand the dynamics of specific events over time. Recent work has leveraged these benefits of ONA to study affect dynamics, the shifts in emotions over time, theorizing differences that might be driven by both motivational factors and even learning experiences. This study uses ONA to investigate these relationships within the context of the affective transitions made by students of four behavioral archetypes—*Scanners*, *Worksheet Users*, *Conversers*, and *Roamers*—in Crystal Island, a game designed to support inquiry-based learning in microbiology. Results indicate that while both *Scanners* and *Worksheet Users* exhibited more positively valenced affect dynamics, only *Worksheet Users* showed measurable learning gains—likely due to their greater engagement in systematic exploration and reflection, particularly while experiencing and regulating confusion and frustration. In contrast, *Conversers* and *Roamers* experienced persistent boredom and prolonged confusion, accompanied by low motivation and poor performance. These findings suggest that successful learning in game-based environments depends not only on positive affect but also on the presence of productive struggle and effective emotional regulation, demonstrating the potential of an analytical approach centered around ONA to elucidate the relationships between behavior and affect during learning.

Keywords: Affect Dynamics, Ordered Network Analysis, Player Archetypes, Game-based learning, Educational Games.

1 Introduction

Affect dynamics can shape students' learning experiences. Prior research has shown that affect within digital learning environments is correlated with a range of important constructs, including self-efficacy [1], analytical reasoning [2], learning outcomes [3, 4], and even college enrollment years later [5]. However, most of this work has focused on correlations between the overall proportion of an affective state during learning and single outcomes or motivational measures. This approach often overlooks two critical aspects of affect dynamics: the role of transitions between emotional states—particularly how students resolve negatively-valenced states—and the specific behavioral patterns associated with these dynamics.

Some affective states—such as engaged concentration and boredom—have consistently been associated with positive and negative learning outcomes, respectively (see review in [3]). However, other states typically considered negative, such as confusion and frustration, have shown mixed effects [3]. Some researchers suggest that a moderate amount of these emotions is actually beneficial (e.g., [4]), while others argue for multiple types of frustration and confusion with different impacts on learning [6, 7].

Another perspective is that some students manage these emotions more positively than others—reframing difficulty as a challenge rather than an insurmountable obstacle—which may mediate their impact on learning. The dynamics of how and when affect occurs also matter for outcomes. For example, Liu et al., [4] showed pretest scores and learning gains were positively correlated with confusion and frustration that were distributed across the learning experience (as opposed to experiencing longer episodes).

Recently, researchers have used Epistemic Network Analysis (ENA [8]) and Ordered Network Analysis (ONA [9]) to investigate these issues [10, 11]. In particular, ONA difference models have been applied to compare affect dynamics among students with different levels of situational interest and learning gains [10], and to track how these dynamics change over time in relation to students' self-efficacy—yielding insights that have advanced affective theory [11].

This paper builds on that work by using ONA to connect affective and behavioral patterns in Crystal Island, an open-world educational game designed to foster interest and support inquiry-based learning in microbiology [12]. Specifically, we investigate the main differences in the affect dynamics of four player/behavioral archetypes—*Scanners*, *Worksheet Users*, *Conversers*, and *Roamers*—previously identified through clustering and ONA in earlier research [13]. By studying these differences, we aim to deepen our understanding of how affect dynamics relate to students' behavioral patterns and, ultimately, their learning outcomes and motivation. Such insights can inform the design of more targeted and effective interventions that respond not only to students' observed behaviors or affective states in isolation, but to the interplay between the two.

2 Related Work

D'Mello and Graesser's [14] highly influential Theory of Affect Dynamics describes how students transition between multiple epistemic emotions (e.g., engaged concentration ↔ confusion ↔ frustration ↔ boredom). Drawing from Control-Value Theory [15], this model proposes that learning is most likely when an impasse is detected and successfully resolved. However, if the impasse remains unresolved, students may transition into frustration and eventually boredom, which can hinder learning.

D'Mello and Graesser's [14] model has inspired numerous empirical studies. However, its use of the L-statistic [16] for analyzing affective transitions may have led researchers to ignore other important affective transitions since it normalizes transition probabilities based on the base rate of each affective state, simultaneously inflating the frequency of transitions involving rare affective states [17], and obscuring transitions involving more common ones [18]. Concerns have also been raised about how the L-

statistic handles self-transition [17, 18], which could be important for improving our understanding of the duration of various affective states.

Unlike the base-rate normalization employed by the L-statistic, ENA applies cosine normalization across each student's affective transitions [8], including self-transitions in the case of ONA [9], emerging as alternatives to address L-statistic's limitations [10, 11, 18]. This approach highlights the most prevalent transitions within students' affective sequences rather than amplifying those that appear notable due to low base rates.

Following this methodological shift, Ocumpaugh et al. [11] used ONA to analyze affective labels obtained through machine learning-based detectors to develop the Skills, Difficulty, Value, Efficacy, and Time (SDVET) model. The model proposes that boredom is most likely when a student's skill level greatly exceeds the task's difficulty. Boredom or canonical (unpleasant) frustration can occur when the task is far more difficult than the student's skill level. Canonical frustration may also arise when skill and difficulty are matched, but the student places low value on the task. Positive affective states, including Gee's [7] pleasurable frustration, emerge when skill and difficulty are well matched and task value is high or when a high self-efficacy student faces a task slightly above their current skill level and feels momentarily confused. However, even when the perceived value of a task is high, students with low self-efficacy may experience intolerable confusion if the task greatly exceeds their perceived skill level, increasing the likelihood of transitioning to boredom or canonical frustration.

Although ENA and ONA have been applied to investigate affect dynamics for students with different motivational traits or levels of learning gains [10, 11] and have also been applied to analyze action sequences in game-based learning environments [19, 20], research has not yet used these methods to connect specific behavioral patterns to sequences of student affect. Instead, previous studies have used regression and correlation analyses to examine associations between aggregated affective states and individual actions within educational games (e.g., [21, 22]). These approaches treat the data as isolated events, rather than as sequences of strategic behaviors or affect. In this study, we use ONA to analyze how affect dynamics relate to specific behavioral patterns—an important next step for identifying the behavioral profiles that may underlie or co-occur with specific affective states and transitions.

3 Methods

3.1 Context and Participants

This study analyzes data from 122 middle school students in the southeastern United States played the educational game Crystal Island as part of their standard science curriculum. The group was gender-balanced (52 male, 65 female, and 5 self-describing in another way). The sample reflected substantial demographic diversity, including historically underrepresented groups (43% Black, 24% Latinx, 5% Asian, and 4% from other racial backgrounds).

Crystal Island is a single-player, open-world game designed to promote interest and support inquiry-based learning in microbiology [12]. Players adopt the role of investigative scientists, diagnosing the pathogen causing a disease outbreak at a remote

research station and identifying its mode of transmission. Gameplay begins in a tutorial area where students learn game mechanics, including how to interact with non-player characters (NPCs), collect and analyze objects, read scientific texts, and complete concept matrices summarizing findings. Upon completing the tutorial, players are directed to the infirmary for additional guidance but are free to explore the island at will. To test hypotheses about possible viral or bacterial agents, players collect and scan items in a virtual laboratory and record their findings in a digital worksheet designed to encourage reflection and support more strategic decision-making in future gameplay.

First, students completed a series of established, validated surveys including a science knowledge pre-test (scored from 0 to 17), and Britner & Pajares's [23] Self-Efficacy survey and Linnenbrink-Garcia et al.'s [24] Situational Interest survey. Next, students engaged with Crystal Island over two class periods. Upon completing the game, they responded to a post-test (identical to the pre-test) to measure knowledge gains and multiple motivational scales designed to assess student perceptions of the game, including five subscales from the Intrinsic Motivation Inventory [25].

3.2 Player Archetypes Identification

The player archetypes in this study were first identified in [13] using ONA and clustering on behavioral data from Crystal Island. In that study, we identified several behavioral codes within students' logs (e.g., interacting with an NPC) to generate an ONA model with the students as the unit variable. We then applied K-means clustering to the model's connection weights (CW), revealing four distinct player archetypes (see [13] for further details).

Table 1. Knowledge assessment and motivational measures across the four behavioral archetypes. Standard deviations are shown in parentheses. Significant differences are shown in bold. Table adapted from Zambrano et al. [13].

Instrument	Timing	Scanners	Worksheet	Conversers	Roamers
N Students	Pre-	29	25	50	18
	Post-	12	12	21	3
Knowledge Scores	Pre-	6.58 (2.42)	5.92 (2.34)	6.02 (2.80)	5.44 (2.17)
	Post-	**6.42 (4.25)**	**6.83 (2.92)**	**4.81 (1.72)**	**4.33 (2.17)**
Linnenbrink-Garcia's SI	**Pre**	**4.21 (1.11)**	**4.47 (1.17)**	**3.73 (1.01)**	**3.37 (1.00)**
Britner & Pajares' SE	**Pre**	**3.99 (0.87)**	**4.11 (0.87)**	**3.65 (1.16)**	**3.40 (1.15)**
Value-utility (IMI)	Post	4.12 (2.06)	3.44 (1.36)	3.44 (1.65)	2.57 (1.26)
Effort-importance (IMI)	Post	4.15 (1.37)	3.98 (0.86)	3.66 (1.42)	2.96 (0.84)
Pressure-tension (IMI)	Post	3.67 (1.17)	3.44 (0.88)	3.35 (1.16)	3.20 (0.81)
Interest-enjoyment (IMI)	**Post**	**4.26 (1.02)**	**4.14 (1.18)**	**3.31 (1.29)**	**3.09 (0.58)**
Perceived Competence (IMI)	**Post**	**4.36 (1.81)**	**4.01 (1.23)**	**3.41 (1.36)**	**2.87 (0.74)**

These archetypes included two groups who appear disengaged from science-related activities—*Roamers* and *Conversers*—and two with stronger engagement—*Scanners* and *Worksheet Users*. *Roamers* primarily wandered through the game environment without interacting with key elements, while *Conversers* engaged in extended dialogues

with NPCs but neglected core science tasks. In contrast, *Scanners* actively tested hypotheses to solve the game's central mystery, but failed to systematize their findings and relied on a more trial-and-error approach, while *Worksheet Users* tested fewer hypotheses but systematically recorded their findings throughout gameplay. These archetypes not only differed in gameplay behaviors but also showed statistically significant differences in situational interest, self-efficacy, perceived enjoyment, and post-test performance. The science-engaged groups—*Scanners* and *Worksheet Users*—achieved the highest motivational scores, and notably, the *Worksheet Users* were the only group who achieved positive learning gains (Table 1).

3.3 Coding Affective States

The affect codes used in this study are based on previously published interaction-based affect detectors [26], which model what students do within a system when experiencing specific emotions, allowing researchers to infer affective states in real time (see [27]). These detectors were developed using labels derived from (a) in-game self-reports (SR) and (b) BROMP-based classroom observations. For the SR-ground truth, students identified one of six affective states: boredom, focus, confusion, happiness, frustration, or nervousness [26]. BROMP observations [28] included boredom, engaged concentration, confusion, delight, and frustration. Nervousness was excluded, as it is difficult to observe directly. To align the SR labels with the standard BROMP framework, child-friendly terminology was used—for example, focus for engaged concentration and happiness for delight.

The SR- and BROMP-based detectors were developed independently. Each data source was used to train a distinct suite of machine learning–based affect detectors, which were cross-validated to ensure generalizability to new student populations. Following prior work using BROMP data (see [26, 27]), both detector suites were trained using 20-second clips of students' interaction logs. Each suite was then applied and analyzed separately, assigning one label per clip from each detector (the SR-based and the BROMP-based) with the highest predicted probability. Prior research suggests that SR- and BROMP-based detectors capture complementary aspects of students' affective states [26], making the differences between them potentially meaningful.

3.4 Ordered Network Analysis and Affect Dynamics

Building on prior work that used ENA and ONA to examine affective transitions in game-based learning environments [10, 11, 18], we use ONA [9] to analyze the affect sequences of the four player archetypes identified in [13]. ONA was chosen for its ability to capture both directionality and self-transitions between affective states. Each line of data corresponds to a 20-second segment of gameplay and includes two labels—one from each suite of affect detectors. Inter-rater reliability was not computed because these labels were assigned by detectors already shown to agree with human coders or self-report, on data not used to develop the detectors.

The unit and conversation variables are defined as the full gameplay session (mean=41.6 min., SD=15.6), since there were no natural breaks during the session. A

moving stanza window of size two is used, so that only transitions between consecutive lines (i.e., affective states) are considered—a standard approach in affective transition analysis [10, 11, 18]. Students were grouped into the four identified archetypes: *Roamers*, *Conversers*, *Scanners*, and *Worksheet Users*.

A Kruskal–Wallis test was conducted to assess whether individual connection weights in the epistemic network models—representing transitions between different affective states or persistence within the same affective state—differ significantly across the four archetypes. These weights were extracted from the WebENA tool [29]. We applied the Benjamini–Hochberg (BH) procedure [30] to control for the false discovery rate resulting from multiple comparisons. Additionally, a Monte Carlo (MC) analysis [31] with 10,000 iterations was employed to determine whether the overall number of significant results was more likely than chance.

4 Results

4.1 Affect States Across Archetypes

This study examines significant differences across four archetypes based on BROMP- and SR-based detectors and ONA. Table 2 presents the distribution of affective states detected using both suites of detectors for each behavioral group. Overall, notable differences in the base rates of affective states can be observed between the two ground truth sources—a finding that aligns with previous analyses [32]. Statistically significant differences across groups occur for five affective states-ground truth pairs after applying the BH correction. Within the suite of BROMP-based detectors, boredom, delight, and confusion show significant differences, while for SR-based detectors, only boredom and concentration show differences. Distributional differences emerged between students who were more engaged in scientific behaviors and showed the highest motivational measures (*Scanners* and *Worksheet Users*) and those who were less engaged (*Conversers* and *Roamers*).

Table 2. Comparison of affect detection across the four behavioral archetypes for each ground truth label (BROMP and SR). Boldface shows statistically significant differences after applying BH correction.

	Scanners		**Worksheet**		**Conversers**		**Roamers**		**p-values**	
	BROMP	**SR**	**BROMP**	**SR**	**BROMP**	**SR**	**BROMP**	**SR**	**BROMP**	**SR**
Concent.	89%	**37%**	85%	**34%**	84%	**27%**	76%	**15%**	0.314	**<0.001**
Boredom	**0%**	**21%**	**2%**	**25%**	**7%**	**39%**	**7%**	**56%**	**0.005**	**<0.001**
Del./Happ.	**3%**	23%	**1%**	16%	**0%**	12%	**0%**	9%	**0.007**	0.350
Confusion	**3%**	13%	**5%**	15%	**6%**	12%	**15%**	11%	**0.008**	0.483
Frustration	5%	5%	7%	9%	3%	9%	2%	9%	0.101	0.261
Nervous		2%		2%		1%		1%		0.763

In the BROMP-based detection, delight is observed more frequently among *Scanners* and *Worksheet Users* (over 1% for both) than among *Conversers* and *Roamers* (less than 0.5% for both). Boredom, by contrast, was significantly more frequent among

Conversers and *Roamers* (7%), compared to 2% for *Worksheet Users* and no instances for *Scanners*. Significantly higher rates of confusion were also observed among *Conversers* and *Roamers* (6% and 15%, respectively), suggesting that these students, mainly *Roamers*, may be struggling. *Scanners* and *Worksheet Users* only experienced confusion in less than 5% of their gameplay.

SR-based results also revealed similar patterns. For example, while self-reported boredom is relatively more frequent than observed boredom, both *Scanners* and *Worksheet Users* report lower boredom rates (20% and 25%, respectively) than *Conversers* and *Roamers* (39% and 56%) as observed according to the BROMP-based detection. Similarly, *Scanners* and *Worksheet Users* exhibit significantly higher concentration rates (37% and 34%, respectively) than *Conversers* and *Roamers* (26% and 15%), although self-reported engaged concentration was much less common than for BROMP observations, where engaged concentration remains above 75% across all students—a pattern commonly reported in BROMP studies (e,g, [33]). No significant group differences were found for frustration or nervousness across both suites of detectors—although notably, *Worksheet Users*, the group with the highest learning gains, also showed the highest frustration rate according to the BROMP-based detectors and the second highest according to the SR-based detectors, a finding we consider later in the context of frustration-related transitions.

4.2 Affect Sequences Detectors

Figure 1 compares the affect dynamics experienced by the four player archetypes based on BROMP-based detectors and ONA models. Using the Kruskal-Wallis test, we identified ten transitions with $p < 0.05$—substantially more than the maximum of four transitions expected by chance according to MC simulations (95% CI: 0–4). As shown in Table 3, these include six cross-affect transitions (Conc → Conf, Frus → Conc, Del → Conc, Conc → Del, Conc → Bor, and Del → Conf) and four of the five self-transitions (sustained frustration, confusion, boredom, and delight). Although only one transition (Conc → Del) remained significant after applying the BH correction, the MC results suggest that the overall pattern of findings is unlikely to be due to chance—some of these additional findings are likely to be genuine but we cannot be certain as to which.

Likewise, the ONA models based on SR-based detectors for the same archetypes are shown in Figure 2. Only two transitions were statistically significant after applying the BH correction (sustained concentration and boredom; Table 4). Two others (Conf → Del, Del → Conf) initially appeared significant but did not hold after the BH correction. However, these transitions mirror patterns observed in the BROMP-based models, indicating they may still represent meaningful group-level differences even if they remain tentative and uncertain.

The primary source of group differences across both suites of detectors involves recurrent experiences of boredom (i.e., Bor → Bor). According to the SR detectors, *Worksheet Users* and *Scanners* show lower levels of sustained boredom (CW=0.384 and 0.320, respectively) compared to *Roamers* and *Conversers* (CW=0.785 and 0.592; $p<0.001$). The BROMP detectors reveal a similar pattern. Recurrent boredom is highest

among *Roamers* (CW=0.106) and *Conversers* (CW=0.095), whereas *Scanners* do not exhibit a single instance of boredom (p=0.014).

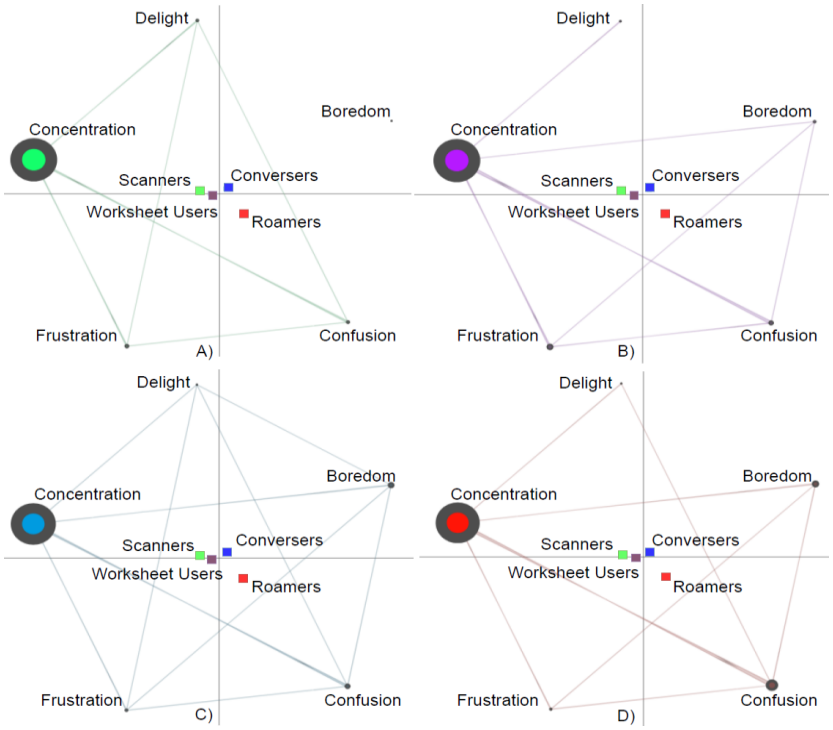


Fig. 1. ONA models of BROMP-based affective transitions for the four behavioral archetypes. A) Scanners (green); B) Worksheet Users (purple); C) Conversers (blue); D) Roamers (red).

Table 3. BROMP-based self-transitions and transitions with CWs showing differences across the four archetypes at $p < 0.05$. Transitions shown in bold are statistically significant after applying the BH correction.

Transition	Scanners (N=29)	Worksheet (N=25)	Conversers (N=50)	Roamers (N=18)	p-val
Concentration → Concentration	0.992	0.965	0.941	0.907	0.271
Frustration → Frustration	0.050	0.090	0.030	0.018	0.037
Confusion → Confusion	0.026	0.048	0.064	0.198	0.007
Boredom → Boredom	0.000	0.030	0.095	0.106	0.014
Delight → Delight	0.028	0.011	0.004	0.003	0.021
Concentration → Confusion	0.013	0.021	0.015	0.026	0.030
Frustration → Concentration	0.007	0.009	0.002	0.004	0.016
Delight → Concentration	0.004	0.003	0.001	0.001	0.033
Concentration → Delight	**0.007**	**0.003**	**0.001**	**0.002**	**<0.001**
Concentration → Boredom	0.000	0.001	0.003	0.005	0.005
Delight → Confusion	0.001	0.000	0.000	0.001	0.013

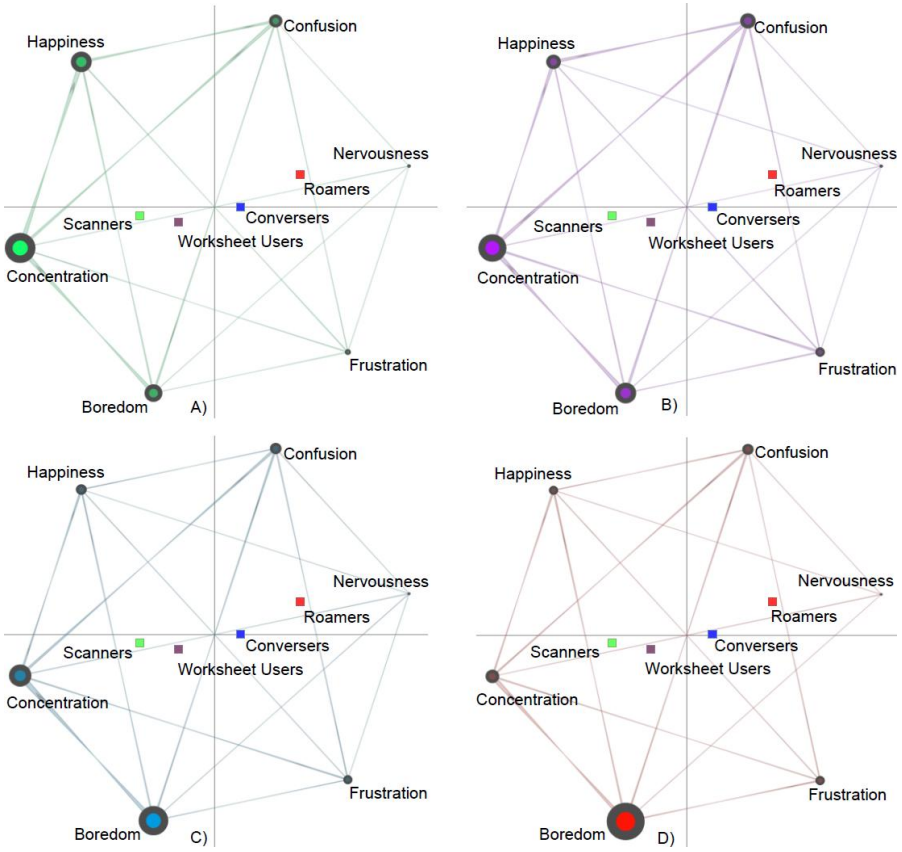


Fig. 2. ONA models of SR-based affective transitions for the four behavioral archetypes. A) Scanners (green); B) Worksheet Users (Purple); C) Conversers (Blue); D) Roamers (Red).

Table 4. SR-based self-transitions and transitions with CWs showing differences across the four archetypes at $p < 0.05$. Transitions shown in bold are statistically significant after applying the BH correction.

Transition	Scanners (N=29)	Worksheet (N=25)	Conversers (N=50)	Roamers (N=18)	p-val
Concentration → Concentration	**0.573**	**0.518**	**0.395**	**0.213**	**<0.001**
Boredom → Boredom	**0.320**	**0.384**	**0.592**	**0.785**	**<0.001**
Confusion → Confusion	0.210	0.245	0.179	0.182	0.580
Happiness → Happiness	0.372	0.242	0.177	0.135	0.178
Frustration → Frustration	0.071	0.138	0.142	0.148	0.358
Nervousness → Nervousness	0.028	0.025	0.013	0.008	0.746
Confusion → Happiness	0.015	0.021	0.006	0.005	0.020
Happiness → Confusion	0.016	0.017	0.004	0.007	0.016

Another important difference involved recurrent engaged concentration, where these dynamics are reversed. In the SR detector data, *Scanners* and *Worksheet Users*

exhibited significantly higher levels of recurrent concentration (CW=0.573 and 0.518, respectively) compared to *Conversers* and *Roamers* (CW=0.395 and 0.213; p<0.001). *Scanners* and *Worksheet Users* also appear to show higher recurrent concentration (CW=0.992 and 0.965) than *Conversers* and *Roamers* (CW=0.941 and 0.907; p=0.271) within the BROMP data. Although this difference is not statistically significant, this is quite possibly due to the high base rate of engaged concentration in the BROMP data.

These findings suggest that the persistence of boredom—and the disruption of concentration—plays a central role in differentiating the affect dynamics of the four behavioral archetypes. In line with these results, *Roamers* and *Conversers* also show more frequent transitions from concentration to boredom (CW=0.005 and 0.003, respectively) than the other two groups (CW<0.001; p=0.005).

Differences are also seen in confusion and frustration, but only when considering BROMP-based detectors. In the SR data, recurrent confusion and frustration are nearly identical across the four archetypes, and cross-affect transitions are also not significant. However, in the BROMP-based data, both sustained frustration and sustained confusion showed substantial differences across groups, with *Worksheet Users* showing higher sustained frustration than the other three archetypes (CW = 0.090 vs CWs = 0.050, 0.030, 0.018; p=0.037), and with *Roamers* showing much higher rates of sustained confusion than the other archetypes (CW = 0.198 vs CWs = 0.026, 0.048, 0.064; p=0.007). These differences were also observed in several transitions involving confusion and frustration (Conc → Conf, Frus → Conc, Del → Conf).

In the BROMP detectors, additional differences emerge in the transitions involving delight. Recurrent delight was highest among *Scanners* (CW=0.028), and *Worksheet Users* (CW=0.011), while *Conversers* and *Roamers* (CW=0.004 and 0.003) displayed much lower persistence in this positive state (p=0.021). Likewise, the transition from concentration to delight is more likely among *Scanners* and *Worksheet Users* (CW=0.007 and 0.003, respectively) than for *Conversers* and *Roamers* (CW=0.001 and 0.002; p<0.001).

5 Discussion and Conclusion

5.1 Understanding Associations between Affect and Behavioral Dynamics

This study advances our understanding of how affect dynamics interact with behavioral patterns, motivation, and learning outcomes in game-based learning environments, extending current models of affect dynamics [11, 14]. By applying ONA and statistical analyses to player behavior, we identify significant differences in the affect dynamics of four distinct behavioral archetypes within the game Crystal Island. Among these groups, *Scanners* (and, to a lesser extent, *Worksheet Users*) appeared to have a more positively valenced affective experience, persisting longer in positive states such as engaged concentration, happiness, and delight. This pattern of positive affective experiences aligns with previous analyses of external motivational measures in this data (Table 1). For example, Zambrano et al. [34] showed that higher motivational measures (situational interest and self-efficacy) are associated with longer episodes of engaged concentration, greater happiness, and more frequent (but shorter) episodes of delight.

These findings also align with the SDVET framework [11], which proposes that students with high self-efficacy and perceived value—and for whom skill and challenge are well-matched—are more likely to experience positive affective states.

Although *Scanners* exhibited the highest overall post-test motivational measures and the greatest persistence in engaged concentration, happiness, and delight, their post-test scores were similar to—or slightly lower than—their pre-test scores. In contrast, *Worksheet Users* experienced higher levels of confusion and frustration than *Scanners* but also demonstrated greater self-efficacy, which, according to the SDVET model, enabled them to regulate these negative emotions better. They showed stronger transitions from confusion and frustration to concentration and happiness, and were the only group to achieve learning gains between the pre- and post-tests.

Understanding the relationships between behavioral archetypes and their affective patterns is not always straightforward. *Scanners* reported the highest levels of perceived competence, had the highest pre-test scores, and exhibited the lowest levels of confusion in the BROMP-based data. It is plausible that some of these students were overconfident in their trial-and-error strategy—especially when compared to *Worksheet Users*, who, experiencing more confusion and frustration, were more systematic in organizing and documenting their findings. This overconfidence and reduced confusion and frustration could have hindered deeper learning. However, confusion patterns for *Scanners* are complex, as they exhibited higher (though not significant) rates of sustained confusion in the SR-based data compared to *Roamers* and *Conversers*, but still lower than *Worksheet Users*.

Overall, these findings highlight the potential benefits of manageable levels of confusion and frustration (cf. [35]). As seen in the affect dynamics of *Worksheet Users*, when these states are not excessive and are effectively managed—particularly by students with sufficient self-efficacy or the ability to engage in self-regulatory behaviors such as externalizing, systematically organizing, and reflecting on their findings (behaviors not yet fully integrated into affective models)—they can positively support learning. However, to fully establish the connection between affect dynamics and specific in-game behaviors, future research is still needed to understand whether confusion primarily emerged in response to failure (e.g., among *Scanners* after scanning the wrong object or submitting an incorrect solution), and whether, for students with stronger self-regulatory skills (e.g., *Worksheet Users*), confusion may play a more generative role—prompting desirable scientific behaviors such as reflection and the systematic organization of findings during gameplay. Similarly, it is worth investigating whether the notably high levels of frustration observed in *Worksheet Users* reflect canonical forms of frustration following failure, or instead a more productive (or even pleasurable; [7]) form of frustration in which students remain engaged, enjoy the challenge, and are further motivated to reflect and organize their findings.

5.2 Proposing Interventions for Educational Games

This study identifies multiple types of players in educational games, helping researchers and designers to tailor interventions to their needs. Some students (e.g., *Scanners*) may have the prior knowledge, initial interest, and self-efficacy to engage with and

enjoy the game, but may not experience enough confusion or frustration to trigger strategy revision and support deeper learning.

Incorporating elements that prompt students to question their current gameplay approach and reflect on its effectiveness—such as in-game questions that challenge their apparent trial-and-error strategy or their prior domain-specific knowledge—could introduce mild, productive confusion (cf. [35]), encourage the use of self-regulation tools (e.g., an in-game worksheet), and promote more systematic exploration. This could involve not only testing hypotheses to solve the game's mystery but also engaging more deeply with domain-specific content through interactions with NPCs or gathering additional information beyond direct experimentation to address broader scientific questions posed by the game. Given their already high post-gameplay interest and positive affect dynamics, these students are likely to respond constructively to such prompts, viewing them as cognitive challenges rather than emotional burdens. Similarly, facilitators could introduce brief, targeted moments of confusion through in-person questioning followed by guided support, stimulating curiosity and reflection without overwhelming the learner. By fostering this kind of engagement, these students may become more aware of the limitations of their current strategies or knowledge and adopt more structured approaches that are more strongly associated with learning gains.

In contrast, other students (e.g., *Roamers*) may lack the foundational knowledge necessary to understand the game's mechanics and learn from the experience or perceive little value in its content. As suggested by our results and theoretical models [11], these students tend to show high levels of boredom and confusion, along with low motivation and limited foundational knowledge. All these factors can hinder meaningful engagement and their ability to regulate negative emotions, reinforcing a negative motivational outlook toward the game and its subject matter, as reflected in the motivational surveys.

Interventions for these students should focus on scaffolding subject knowledge and game mechanics to reduce the mismatch between skill and challenge, aiming to enhance their engagement. This support could be integrated into the tutorial or early stages of the game by providing more detailed information about the task, the available tools, and relevant domain-specific concepts—especially when students display behaviors associated with confusion, such as wandering or roaming without completing expected actions. To address potential boredom, interventions should also account for causes beyond unresolved confusion, misunderstandings, or lack of skill (as theorized by D'Mello & Graesser [14]), such as low interest, motivation, or perceived value—as manifested by direct transitions from concentration to boredom without passing through confusion. In such cases, both the game (e.g., via pop-up messages) and the facilitator can help students recognize the educational value of the experience and the relevance of its content. These strategies can boost situational interest to a level that supports sustained engagement and prevents the prolonged boredom that undermines learning.

Similar interventions could be designed for students with higher prior knowledge and slightly stronger motivational measures (e.g., *Conversers*), who nonetheless show affect dynamics and motivational profiles similar to the most disengaged group. These students may still engage with certain game mechanics (e.g., interacting with NPCs), which are important to identify and leverage for delivering targeted messages that address motivational, affective, or cognitive challenges. Doing so can help support

students as they navigate difficulties and reduce the risk of boredom from repeated struggles. Implementing such interventions early in the game could prevent the development of a negative perception of the game, at which point interventions might become less effective.

In conclusion, this study demonstrates the value of ONA to investigate how affect dynamics in educational games relate to students' behavioral patterns, motivation, and learning outcomes. The contrast between *Scanners* and *Worksheet Users* highlights the importance of productive emotional struggle (particularly confusion and frustration) as a potential driver of deeper learning when effectively regulated. In contrast, the high levels of unresolved confusion and boredom observed among *Conversers* and *Roamers* show how mismatches between skill and challenge, low perceived value, and insufficient support can lead to persistent negative affect and disengaged gameplay, ultimately hindering learning. These findings have important implications for affective research and the design of adaptive learning environments. Effective interventions should consider students' cognitive and motivational characteristics, real-time affective states, and in-game behaviors. By aligning gameplay mechanics, scaffolds, and feedback with students' evolving emotional and behavioral profiles, designers and educators can better foster both engagement and meaningful learning across diverse learner types.

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.

Disclosure of Interests. The authors have no competing interests to declare that are relevant to the content of this article.

References

1. McQuiggan, S.W., Lester, J.C.: Modelling affect expression and recognition in an interactive learning environment. International Journal of Learning Technology. 4, 216–233 (2009). doi: 10.1504/IJLT.2009.028807.
2. D'Mello, S., Person, N., Lehman, B.: Antecedent-consequent relationships and cyclical patterns between affective states and problem solving outcomes. In: Artificial Intelligence in Education. pp. 57–64. Ios Press (2009).
3. Karumbaiah, S., Baker, R., Tao, Y., Liu, Z.: How does Students' Affect in Virtual Learning Relate to Their Outcomes? A Systematic Review Challenging the Positive-Negative Dichotomy. In: 12th International Learning Analytics and Knowledge Conference. pp. 24–33. Association for Computing Machinery, New York, NY, USA (2022).
4. Liu, Z., Pataranutaporn, V., Ocumpaugh, J., Baker, R.: Sequences of frustration and confusion, and learning. In: International Conference on Educational Data Mining. pp. 114–120 (2013).
5. San Pedro, M.O., Baker, R., Bowers, A., Heffernan, N.: Predicting college enrollment from student interaction with an intelligent tutoring system in middle school. In: International Conference on Educational Data Mining (2013). doi: 10.1007/978-3-642-39112-5_5.

6. Baker, R.S., Cloude, E., Andres, J.M.A.L., Wei, Z.: The Confrustion Constellation: A New Way of Looking at Confusion and Frustration. Cognitive Science. 49, e70035 (2025). doi: 10.1111/cogs.70035.
7. Gee, J.P.: Situated language and learning: A critique of traditional schooling. Routledge (2012). doi: 10.4324/9780203594216.
8. Shaffer, D.W., Collier, W., Ruis, A.R.: A tutorial on epistemic network analysis: Analyzing the structure of connections in cognitive, social, and interaction data. Journal of Learning Analytics. 3, 9–45 (2016). doi: 10.18608/jla.2016.33.3.
9. Tan, Y., Ruis, A.R., Marquart, C., Cai, Z., Knowles, M.A., Shaffer, D.W.: Ordered network analysis. In: International Conference on Quantitative Ethnography. pp. 101–116. Springer (2022). doi: 10.1007/978-3-031-31726-2_8.
10. Nasiar, N., Zambrano, A.F., Ocumpaugh, J., Goslen, A., Rowe, J., Vandenberg, J., Esiason, J., Hutt, S.: The Influence of Different Measurement Approaches on Student Affect Transitions Using Ordered Networks. In: Kim, Y.J. and Swiecki, Z. (eds.) Advances in Quantitative Ethnography. pp. 195–203. Springer Nature Switzerland (2024).
11. Ocumpaugh, J., Nasiar, N., Zambrano, A.F., Goslen, A., Vandenberg, J., Esiason, J., Rowe, J., Hutt, S.: Refocusing the lens through which we view affect dynamics: The Skills, Difficulty, Value, Efficacy and Time Model. Presented at the Learning Analytics and Knowledge Conference , Dublin, Ireland (2025). doi: 10.1145/3706468.3706495.
12. Rowe, J.P., Shores, L.R., Mott, B.W., Lester, J.C.: Integrating Learning, Problem Solving, and Engagement in Narrative-Centered Learning Environments. International Journal of Artificial Intelligence in Education. 21, 115–133 (2011). doi: 10.3233/JAI-2011-019.
13. Zambrano, A.F., Barany, A., Ocumpaugh, J., Nasiar, N., Vandenberg, J., Goslen, A., Esiason, J., Rowe, J., Hutt, S.: 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. (2025). doi: 10.1007/s10956-025-10213-4.
14. D'Mello, S., Graesser, A.: Dynamics of affective states during complex learning. Learning and Instruction. 22, 145–157 (2012). doi: 10.1016/j.learninstruc.2011.10.001.
15. Pekrun, R.: The Control-Value Theory of Achievement Emotions: Assumptions, Corollaries, and Implications for Educational Research and Practice. Educational Psychology Review. 18, 315–341 (2006). doi: 10.1007/s10648-006-9029-9.
16. D'Mello, S., Taylor, R. S., Graesser, A. C: Monitoring affective trajectories during complex learning. In: Proceedings of the annual meeting of the cognitive science society (2007).
17. Karumbaiah, S., Baker, R.S., Ocumpaugh, J., Andres, J.Ma.A.L.: A Re-Analysis and Synthesis of Data on Affect Dynamics in Learning. IEEE Transactions on Affective Computing. 14, 1696–1710 (2023). doi: 10.1109/TAFFC.2021.3086118.
18. Karumbaiah, S., Baker, R.S.: Studying affect dynamics using epistemic networks. In: Advances in Quantitative Ethnography: Second International Conference, ICQE 2020, Malibu, CA, USA, February 1-3, 2021, Proceedings 2. pp. 362–374. Springer (2021).
19. Scianna, J., Gagnon, D., Knowles, B.: Counting the Game: Visualizing Changes in Play by Incorporating Game Events. In: Advances in Quantitative Ethnography. pp. 218–231. Springer International Publishing (2021). doi: 10.1007/978-3-030-67788-6_15.
20. Zambrano, A.F., Barany, A., Ocumpaugh, J., Nasiar, N., Hutt, S., Goslen, A., Rowe, J., Lester, J., Wiebe, E., Mott, B.: Cracking the code of learning gains: Using Ordered Network

Analysis to Understand the Influence of Prior Knowledge. In: International Conference on Quantitative Ethnography. pp. 18–33. Springer (2023). doi: 10.1007/978-3-031-47014-1_2.
21. Sabourin, J.L., Lester, J.C.: Affect and engagement in Game-Based Learning environments. IEEE transactions on affective computing. 5, 45–56 (2013). doi: 10.1109/T-AFFC.2013.27
22. Taub, M., Sawyer, R., Lester, J., Azevedo, R.: The Impact of Contextualized Emotions on Self-Regulated Learning and Scientific Reasoning during Learning with a Game-Based Learning Environment. International Journal of Artificial Intelligence in Education. 30, 97–120 (2020). doi: 10.1007/s40593-019-00191-1.
23. Britner, S.L., Pajares, F.: Sources of science self-efficacy beliefs of middle school students. Journal of Research in Science Teaching. 43, 485–499 (2006). doi: 10.1002/tea.20131.
24. Linnenbrink-Garcia, L., Durik, A.M., Conley, A.M., Barron, K.E., Tauer, J.M., Karabenick, S.A., Harackiewicz, J.M.: Measuring situational interest in academic domains. Educational and psychological measurement. 70, 647–671 (2010). doi: 10.1177/0013164409355699.
25. Deci, E.L., Ryan, R.M.: SDT: Questionnaires: Intrinsic motivation inventory (IMI). Retrieved October. 27, 2009 (2007).
26. Zambrano, A.F., Nasiar, N., Ocumpaugh, J., Goslen, A., Zhang, J., Rowe, J., Esiason, J., Vandenberg, J., Hutt, S.: Says Who? How different ground truth measures of emotion impact student affective modeling. In: Proceedings of the 17th International Conference on Educational Data Mining. pp. 211–223. International Educational Data Mining Society, Atlanta, Georgia, USA (2024). doi: 10.5281/zenodo.12729800.
27. Baker, R.Sj., Ocumpaugh, J.: Interaction-Based Affect Detection in Educational Software. The Oxford handbook of affective computing. 233 (2014).
28. Ocumpaugh, J., Baker, R.S., Rodrigo, M.M.T.: Baker Rodrigo Ocumpaugh Monitoring Protocol (BROMP) 2.0 Technical and Training Manual. (2015).
29. Marquart, C.L., Hinojosa, C., Swiecki, Z., Eagan, B., Shaffer, D.W.: Epistemic network analysis (Version 1.7.0) [Software], http://app.epistemicnetwork.org, (2021).
30. Benjamini, Y., Hochberg, Y.: Controlling the false discovery rate: a practical and powerful approach to multiple testing. Journal of the Royal statistical society: series B (Methodological). 57, 289–300 (1995). doi: 10.1111/j.2517-6161.1995.tb02031.x.
31. Metropolis, N., Ulam, S.: The Monte Carlo Method. Journal of the American Statistical Association. 44, 335–341 (1949). doi: 10.1080/01621459.1949.10483310.
32. Baker, R.S.J. d., D'Mello, S.K., Rodrigo, Ma.M.T., Graesser, A.C.: 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, 223–241 (2010).
33. Rodrigo, M.M.T., Baker, R.S.J. d.: Comparing learners' affect while using an intelligent tutor and an educational game. RPTEL. 6, 43–66 (2011). doi: 10.58459/rptel.2011.643-66.
34. Zambrano, A.F., Ocumpaugh, J., Baker, R.S., Vanacore, K., Esiason, J., Vandenberg, J.: 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 (2025).
35. Lehman, B., D'Mello, S., Graesser, A.: Confusion and complex learning during interactions with computer learning environments. The Internet and Higher Education. 15, 184–194 (2012). doi: 10.1016/j.iheduc.2012.01.002.