The Half-Life of Epistemic Emotions: How Motivation Influences Affective Chronometry

Andres Felipe Zambrano[1], Jaclyn Ocumpaugh[1], Ryan S. Baker[1], Kirk Vanacore[1], Jordan Esiason[2], Jessica Vandenberg[2]
[1]University of Pennsylvania
[2]North Carolina State University
{afzambrano97, jlocumpaugh, ryanshaunbaker, kirk.vanacore}@gmail.com,
{jesiaso, jvanden2}@ncsu.edu

ABSTRACT
Research on epistemic emotions has often focused on how students transition between affective states (e.g., affect dynamics). More recently, studies have examined the properties of cases where a student remains in the same affective state over time, finding that the duration of a student's affective state is important for multiple learning outcomes. However, the likelihood of remaining in a given affective state has not been widely studied across different methods or systems. Additionally, the role of motivational factors in the persistence or decay of affective states remains underexplored. This study builds on two prior investigations into the exponential decay of epistemic emotions, expanding the analysis of affective chronometry by incorporating two detection methods based on student self-reports and trained observer labels in a game-based learning environment. We also examine the relationship between motivational measures and affective decay. Our findings indicate that boredom exhibits the slowest decay across both detection methods, while confusion is the least persistent. Furthermore, we found that higher situational interest and self-efficacy are associated with greater persistence in engaged concentration, as identified by both detection methods. This work provides novel insights into how motivational factors shape affective chronometry, contributing to a deeper understanding of the temporal dynamics of epistemic emotions.

Keywords
Affective Chronometry; Affect dynamics; half-life; self-report; observational; motivational differences

1. INTRODUCTION
Research on epistemic emotions has matured over the last decade as researchers have progressed from being able to detect them in the moment (e.g., [3, 32, 35]) to using the resultant detectors to understand patterns in emotion over time and how those patterns relate to learning or motivation (e.g., [1, 16, 22, 23, 26, 37]). However, one aspect that has received less attention in the literature is the temporal dynamics of epistemic emotions—specifically, their duration and decay over time (but see [7, 13]). This includes understanding how long these emotions last, the rate at which they fade, and potential factors such as situational interest or self-efficacy that contribute to a faster decay or extend students' persistence within those affective states.

Andres Felipe Zambrano, Jaclyn Ocumpaugh, Ryan S. Baker, Kirk Vanacore, Jordan Esiason, and Jessica Vandenberg. The Half-Life of Epistemic Emotions: How Motivation Influences Affective Chronometry. In Caitlin Mills, Giora Alexandron, Davide Taibi, Giosuè Lo Bosco, and Luc Paquette (eds.) Proceedings of the 18th International Conference on Educational Data Mining, Palermo, Italy, July, 2025, pp. 317–327. International Educational Data Mining Society (2025).

https://doi.org/10.5281/zenodo.15870173

Understanding typical rates of decay is an important area of research, particularly for affective states like confusion or frustration, where research shows that both too little and too much time in that state is bad for learning [22]. As we are seeking to personalize learning, understanding the timing-related differences in students' affective experience is an important step. In particular, capturing ways in which students' affective experiences are most likely to change over time—and how these changes may be related to learning and motivation measures—can be useful for understanding different tolerances for negatively valenced emotions and/or their antecedents (e.g., problem difficulty, poor connections to student's prior knowledge, etc.).

In this study, we examine differences in students' affective experience within the context of Crystal Island, an online learning system aligned to state curriculum standards for middle school math. In this system, two distinct suites of cross-validated affect detectors—one based on self-reported data and one based on BROMP-observation labels [24]—are already published [37]. We use both sets of detectors—analyzed separately—to examine how student affect changes over time, with a special emphasis on how long a student is likely to persist in each affective state, or its half-life.

Affect dynamic results for the full population are first compared to previous work investigating the half-life of epistemic emotions, namely Botelho et al. [7] and D'Mello et al. [13]. We then look for differences based on the students' motivational measures. Specifically, building on new models of affective states that suggest that student motivational measures may impact their tolerance for difficulty [25], we also examine how the half-life of affective states varies with respect to students' self-efficacy [8] and situational interest [21].

2. LITERATURE REVIEW
2.1 Affective Dynamics
To date, there are a handful of theoretical models related to the ways in which students experience epistemic emotions during learning. One common model is Csikszentmihalyi's Flow Theory [10], which suggests that people experience a state of flow when difficulty of their task is well matched to their skill level. More specific to epistemic emotions is D'Mello and Graesser's [12] affect dynamics model. This model predicts oscillation between engaged concentration and confusion when students are learning, and a path from confusion to frustration to boredom when they are not.

More recently, Ocumpaugh et al. [25] have used Pekrun's Control Value Theory [27] to build upon both Csikszentmihalyi's and

D'Mello and Graesser's frameworks, proposing the Skills, Difficulty, Value, Efficacy, and Timing (SDVET) model. The SDVET model suggests that motivational constructs like self-efficacy and value are important to explain why and when a student might transition from one affective state to another. Much like Csikszentmihalyi's Flow Theory, it hypothesizes that the intersection of skill and difficulty level is important, but it diverges from that model in terms of what emotions are predicted when.

The SDVET model predicts that (a) boredom is likely to occur when students' skill is substantially higher than the task difficulty and that (b) either boredom or canonical frustration are likely when their skill level is well below their difficulty. However, it does not predict that the space between those two areas is entirely occupied by a flow state or engaged concentration. Instead, it hypothesizes that (c) a student whose skill and difficulty are well matched, but who does not value the task, will also experience canonical frustration, whereas (d) a student with matched skill and difficulty, plus high value, will experience the state of flow. Finally, it describes a situation where students with higher self-efficacy are asked to complete tasks with difficulty levels above their current skill level—but within the range where their self-efficacy encourages them to believe they can accomplish this task. Students in this space are predicted to (e) experience intolerable levels of confusion when their value for the task is low—which could make them more likely to transition to boredom or (canonical) frustration. However, when their perceived value of the task is high, they are predicted to experience pleasurable frustration [17]. Notably, the buffering effects of self-efficacy that are predicted in this model (f) are expected to diminish as time persists—meaning that even students with high self-efficacy will not persist in confusion or pleasurable frustration infinitely.

In other words, there are explicit hypotheses about what extended experiences of emotion might indicate, which align both with Ocumpaugh et al.'s [25] empirical data and with the relationships we often see between self-transitions and learning. For example, Nasiar et al., [23] found that extended periods of boredom, confusion, and frustration—all indications that a student was not being appropriately challenged—were associated with low learning gains, while extended periods of engaged concentration and delight were associated with high learning gains. Likewise, Andres et al. [1] showed that sustained boredom was negatively correlated with both pre-test and post-test while sustained delight was positively correlated with post-test and learning gains. None of these studies, however, are able to show what a normal duration of a given affective state might be for a given learner.

2.2 Exponential Decay Research

One analytical approach that can help us to understand typical durations of affective states is the use of exponential decay. To date, there are two primary studies that have looked at the rate of decay of affective states in learning analytics. The first is D'Mello & Graesser's [13] study of the AutoTutor system, in which students and trained human evaluators labeled students' affective states every 20 seconds based on the recorded interactions with the platform. In this study, D'Mello and Graesser observed that both self-labeled and judge-labeled affective states tended to change rapidly within the first minute after students transitioned into the corresponding affect. Specifically, they found a sharp decline in the number of instances where a student still persisted in an affective state recognized one minute earlier. They proposed that this reduction in the probability of persisting in the same affective state could be effectively modeled using exponential decay and introduce the notion of affective half-life, where a half-life represents the point at which a quantity decreases to half of its initial value [7, 13]. This measure provides a more informative perspective than simply calculating the average episode length, as it identifies the point where students are equally likely to remain in or transition out of an affective state.

By modeling these probabilities as exponential decay, D'Mello and Graesser [13] observed that students tend to persist longer in boredom, engaged concentration, and confusion (persistent states), whereas the durations of delight and surprise (transitory states) and frustration (an immediate state) were significantly shorter. Additionally, they found that prior knowledge was negatively correlated with the decay rate of engaged concentration, suggesting that students with higher prior knowledge tend to remain in this affective state for longer. This finding aligns with Csikszentmihalyi's Flow Theory [10], which proposes a balance between challenge and skill level as a key factor in sustaining engagement.

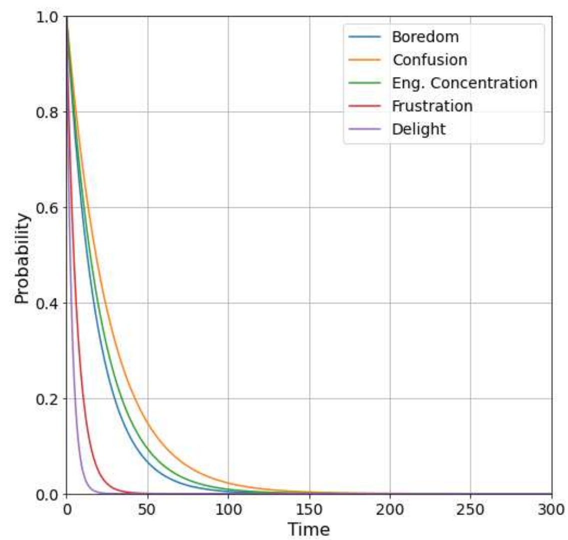

Figure 1. D'Mello & Graesser [13] Exponential Decay Results.

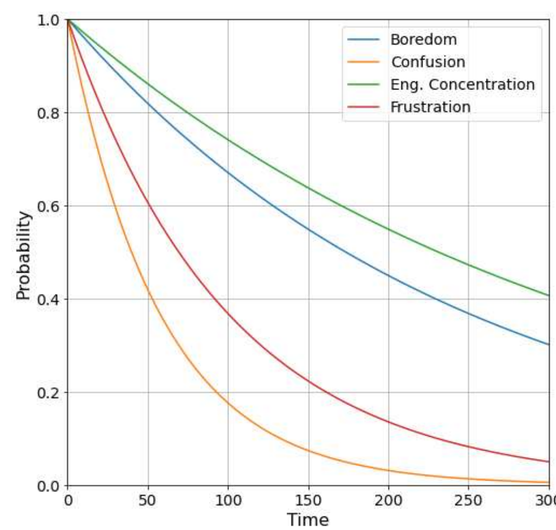

Figure 2. Botelho et al.'s [7] Exponential Decay Results.

Inspired by D'Mello and Graesser's analysis, Botelho et al. [7] studied the same issues within the ASSISTments system [18], where BROMP-based affect detectors made predictions about each student's affective state every 20 seconds. In this study, researchers applied the same exponential function to understand the likelihood that a student would persist in a given affective state once they transitioned into it. However, Botelho et al. [7] found that boredom and engaged concentration often lasted longer than the previously found one-minute window. Consequently, they extended the decay observation period to five minutes to better fit the exponential function and determine the decay rate. The results from D'Mello and Graesser's study are shown in Figure 1, while those from Botelho et al.'s study appear in Figure 2.

3. METHODS

3.1 Data Context

This study analyzes data from an inquiry-based virtual world designed to align with state standards for middle school microbiology. In Crystal Island [30], players act as researchers prompted to identify the cause of an outbreak that has impacted a research team on an island. To complete the game, players navigate multiple locations, interact with non-playable characters (NPCs), collect information from reading materials distributed across the virtual world, and use laboratory tools to test their hypotheses. To support their progress, players are given a concept matrix to organize information from the readings and a worksheet to structure their hypotheses and findings. Figure 3 displays the game interface.

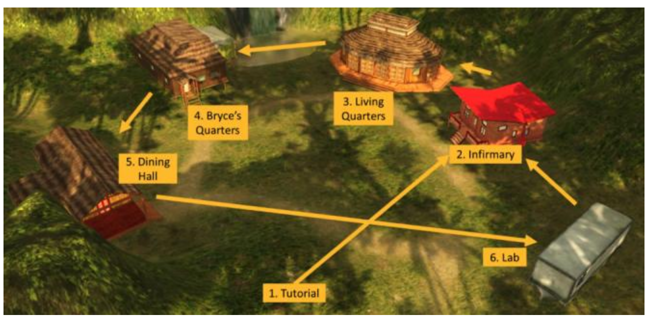

Figure 3. Crystal Island interface with suggested solution path [33].

3.2 Participants

This study analyzes data from 122 middle school students who played Crystal Island at an urban school in the Southeastern United States. The dataset is well-balanced in terms of gender, with 44% of students coming from economically disadvantaged backgrounds, according to school-level statistics. Additionally, over 75% of the participants identify as members of ethnic minority groups, including 43% Black, 24% Latinx, 5% Asian, and 4% from other racial backgrounds. The study was conducted during the students' regular science classes, which lasted approximately one hour per day over a two-day period. All procedures were approved by the Institutional Review Boards (IRBs) of the partner institutions.

3.3 BROMP vs. Self-Report Detectors

This study utilizes interaction-based detectors [3, 6] of epistemic affective states that were previously published and cross-validated [37] to infer their emotions in real-time (see [4, 6] for reviews). These detectors were developed using labels generated from two sources: (a) in-game self-reports (SR) and (b) observations conducted with the BROMP protocol [24].

Data for both of Zambrano et al.'s [37] detectors were collected simultaneously while students were playing the game (average gameplay duration: 41.6 minutes, SD = 15.6). At specific moments during gameplay (e.g., after completing the tutorial, reading three in-game texts, or testing three hypotheses), students were prompted to self-report one of six affective states: boredom (31.7%), focus (28.9%), confusion (13.4%), happiness (16.8%), frustration (7.6%), or nervousness (1.5%). Self-reporting prompts were strategically timed to minimize disruptions to gameplay. In total, 463 self-reports were collected and used to train the self-report-based detectors. No student was prompted to report their affective state more than 10 times throughout the entire gameplay session.

As they were playing, BROMP classroom observations—a momentary time sampling method designed for building detectors of student affect—were conducted. BROMP observers collected data on five affective states: boredom (4.7%), engaged concentration (82.2%), confusion (6.6%), delight (1.3%), and frustration (5.1%). In total, 1,716 observations were collected to train the BROMP-based detectors. Nervousness was excluded from BROMP observations as it is difficult to observe directly. Self-report labels were adapted from BROMP researcher categories to child-friendly language (e.g., "focus" for engaged concentration and "happiness" for delight).

Although there is a sharp contrast in the base rates between the two ground truths (SR vs. BROMP), these differences are consistent with findings from prior studies (e.g., see base rate differences in [5] and [32]). This contrast likely reflects the distinct nature of the signals each method captures and their different limitations (see discussion section, below). Despite their differences, both types of data are associated with multiple outcomes, including learning and motivational measures [37]. Considering both signals can produce a more comprehensive view of students' affective dynamics than either method alone (see discussion in [37]). The BROMP and SR detectors were developed independently of each other, with each set of affective data used to train separate ML-based affect detectors. These were binary one vs all detectors trained using Logistic Regression, Random Forests, and X Gradient Boosting. These detectors were cross-validated at a subgroup level to ensure they could generalize to new populations, and the best-performing model for each affective state (AUC>0.65 for all affective states) was used in this analysis. Consistent with previous work involving BROMP [7], both sets of detectors were applied to 20-second segments of students' log files. Two labels were then assigned to each clip (one from the SR-based detector and one from the BROMP-based detector) based on the highest probability output across categories after adjusting probabilities to account for the base rates of the ground truth in each detector suite, ensuring the distributions matched those observed in the ground truth [23, 25].

3.4 Affective Dynamics and Chronometry

We analyzed affect dynamics using a multistep approach. First, we replicated the affect chronometry approach proposed by D'Mello and Graesser [13] and replicated by Botelho et al. [7]. We segmented each student's sequence of labels into episodes, with each episode representing the continuous duration a student remained in a specific affective state before transitioning to a different one. For example, if the model classified a student as bored for three twenty-second clips before transitioning to frustration for two twenty-second clips and then back to boredom for four twenty-second clips, then they would have experienced three affective episodes—two for boredom and one for frustration.

For each affective state, we used the detector labels to calculate the probabilities of episodes persisting for durations ranging from 20 seconds to 5 minutes in increments of 20 seconds ($Pr(E_t =$

E_{t+20i}), where i represents the number of future clips in which the affective state persists). Note that this is in line with Botelho et al.'s [7] approach, which also used affect detectors, the same granularity, and observed episode lengths similar to those in our study, but represents a deviation from D'Mello & Graesser's approach, as they used the same 20-second increments for labeling but only considered the initial 60 seconds to estimate affective decay. In our study, these calculations produced 16 probabilities (for durations of 0, 20, 40, ..., 300 seconds), which were then used to compute the decay rates and half-lives of each affective state. For this calculation, we ignored the last affective episode of the gameplay because it is not possible to determine what the length of this episode might have been if the student had kept playing.

Unlike previous approaches, we fit the exponential function using a Bayesian regression model, allowing us to calculate not only point estimates but also 95% credible intervals, which represent the probability density distribution within which the true parameter value is likely to fall, given the data and prior knowledge about the parameter (in this case, the prior decay factors found by D'Mello & Graesser, and Botelho et al.). This methodological shift incorporates uncertainty quantification from a different perspective than the more traditional frequentist approach but does not alter the mean estimate, ensuring comparability with previous studies.

In particular, we employed a log regression model to estimate decay rates. The model was specified with an intercept of 1, acknowledging that at time=0, the student is already experiencing the corresponding affective state. Additionally, we used a naïve Gaussian prior (Mean=0, SD=1) to avoid imposing strong initial assumptions about the distribution of decay rates. This choice aligns with findings from D'Mello and Graesser and Botelho et al's studies, who obtained decay rates within the interval (-0.6, 0). Then, using the estimated decay rates λ and their corresponding 95% credible intervals, the half-lives were calculated as $HF = ln(2)/\lambda$.

We compared results across studies and affective states using half-lives instead of decay factors because they are expressed in seconds, offering a straightforward and interpretable measure of the moment when it is more likely to transition to a different affective state rather than still persisting on it. For reference, we include half-lives derived from two types of detectors in D'Mello and Graesser's study—those from external judges' observations and those from participants' self-judgments. In their study, the self-reported data (referred to as self-judgement) was collected retrospectively, with students watching and labeling the video of their own session immediately after ending it. Although this method diverges from our in-the-moment methods, it is the closest prior study to the self-reported data in our analysis.

3.5 Self-Efficacy and Situational Interest

After analyzing affect chronometry across the entire group of students, we examined its association with self-efficacy and situational interest. Prior to playing the game, students completed two external survey measures: Linnenbrink-Garcia et al.'s [21] situational interest scale and Britner & Pajares' [8] self-efficacy scale. Based on their scores for each measure, students were categorized into high, middle, and low groups, considering the middle groups as those students within a standard deviation of the mean of each measure. Students were categorized into high/middle/low separately for each measure. Then, our analysis compared the high and low groups for both variables, in terms of affect chronometry. This categorization method was chosen to reduce the risk of spurious results, which can occur when splitting at the mean, as students near the mean often exhibit similar characteristics.

4. RESULTS

4.1 Exponential Decay (All Students)

Next, we used affective chronometry to analyze the likelihood of students transitioning out of a given affective state and estimate the half-life of each affective state according to both suites of detectors. Figures 4 and 5 present these results for the SR-based detectors and the BROMP-based detectors, respectively. Additionally, Table 1 summarizes the half-life of each affective state, as well as findings from prior studies by Botelho et al. [7] and D'Mello & Graesser [13].

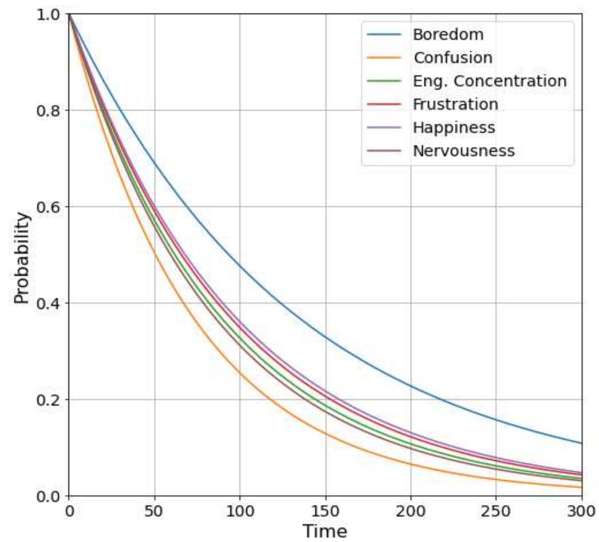

Figure 4. SR-based detectors in Crystal Island.

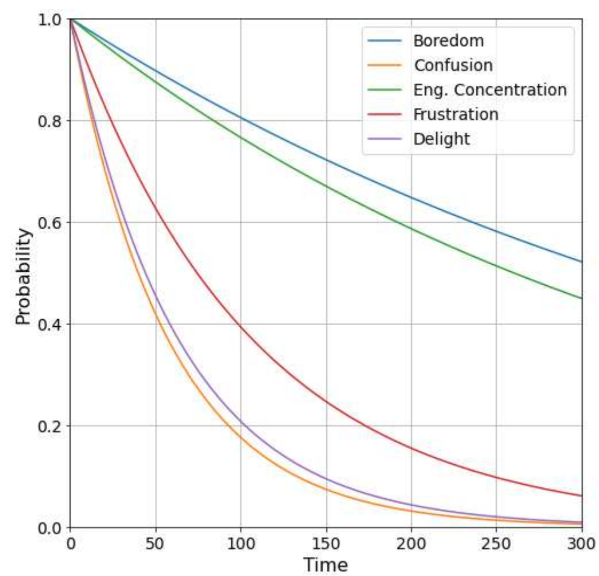

Figure 5. BROMP-based detectors in Crystal Island.

Across both types of detectors, students experienced longer episodes of boredom than in any other affective state, though the difference between the two detectors was still quite large. For the BROMP-based detector, the half-life of boredom was 319.2 seconds (95% CI [296.3, 344.5]), whereas SR-based boredom had a half-life that was nearly two minutes shorter (93.4 sec.; CI [85.5, 101.6]). These findings align with those reported by Botelho et al. (2018), where boredom, as detected using BROMP-based methods, was one of the two affective states with the longest half-life (173.3 sec.). In contrast, D'Mello & Graesser study reported a half-life of less than 15 seconds for boredom for both types of labeling (self-judgment and experienced external judges).

Table 1. Estimated half-lives (in seconds) and 95% Credible Intervals of affective states across different suites of detectors and studies. Results from the current study are shown in bold.

			Credible Intervals	
Affect	**Study**	**Half-life**	**Low**	**High**
Boredom	**SR**	**93.4**	**85.5**	**101.6**
	D'Mello Self	9.6	-	-
	BROMP	**319.2**	**296.3**	**344.5**
	Botelho BROMP	173.3	-	-
	D'Mello Judge	13.9	-	-
Confusion	**SR**	**50.6**	**46.5**	**54.8**
	D'Mello Self	19.8	-	-
	BROMP	**39.9**	**36.1**	**43.8**
	Botelho BROMP	28.8	-	-
	D'Mello Judge	23.5	-	-
Engaged Concentration	**SR**	**61.9**	**52.4**	**71.9**
	D'Mello Self	25.7	-	-
	BROMP	**259.7**	**249.8**	**270.2**
	Botelho BROMP	231.0	-	-
	D'Mello Judge	13.6	-	-
Frustration	**SR**	**65.7**	**54.8**	**77.1**
	D'Mello Self	19.8	-	-
	BROMP	**74.3**	**69.2**	**79.5**
	Botelho BROMP	69.3	-	-
	D'Mello Judge	2.9	-	-
Happiness/ Delight	**SR**	**67.9**	**57.5**	**78.6**
	D'Mello Self	4.1	-	-
	BROMP	**44.1**	**35.0**	**53.4**
	D'Mello Judge	2.0	-	-
Nervousness	**SR**	**59.4**	**55.9**	**63.0**

Engaged concentration showed more differences across detectors. Among our BROMP-based detectors, engaged concentration had the second-longest half-life (259.7 sec., CI [249.8, 270.2]), closely matching the half-life reported in Botelho et al.'s study (231.0 sec.). However, its half-life was more than three minutes shorter when we measured it using the SR-based detectors (61.9 sec., CI [52.4 71.9]), where it ranked in the middle among the other affective states. This value aligns more closely with the shorter half-lives (less than 30 seconds) reported by D'Mello & Graesser (2011). However, in D'Mello & Graesser's study, the half-life of engagement (referred to by them as flow) obtained from external observers (13.6 sec.) was smaller than the half-life obtained from self-labeling (25.7 sec.).

Frustration shows highly consistent results across studies. Among the BROMP-based detectors in this study, it had the third lowest decay rate, with a half-life of just over a minute (74.3 seconds, CI [69.2, 79.5]). Interestingly, the frustration decay curve for the SR-based detectors also revealed a very similar half-life of 65.7 seconds (CI [54.8, 77.1]), suggesting that frustration persists for a comparable duration according to both students (self-reports) and observers (BROMP). These findings are comparable to Botelho et al.'s results (69.3 seconds). Frustration's half-life was much shorter in D'Mello and Graesser's retrospective self-judgments (19.8 seconds). However, as with our BROMP-based detectors, frustration had the third longest half-life. In contrast, their expert judge's frustration half-life value was substantially shorter (2.9 sec.).

The two affective states with the fastest decay rates, according to our BROMP-based detectors, were confusion and delight, with half-lives of 39.9 (CI [36.1, 43.8]) and 44.1 seconds (CI [35.0, 53.4]), respectively. Confusion had a similar half-life of 50.6 (CI [46.5 54.8]) for the SR-based detectors. These findings again closely align with Botelho et al.'s study, where confusion was identified as the affective state with the fastest decay, with a half-life of 28.8 seconds. The authors of that paper did not include delight in their analysis due to its low frequency in that paper's learning system (still, they reported some instances of delight in their data). Similarly, delight was found to have the lowest persistence in the study by D'Mello and Graesser (less than 5 seconds for both labeling methods). However, when analyzing delight through the SR-based detectors, renamed as happiness in this context, the decay rate was slightly slower, with a half-life of 67.9 seconds (CI [57.5, 78.6])—making it the second most persistent state, just below boredom. It is possible that these two constructs do not align as much as originally intended and that students may experience happiness as a more stable, enduring state compared to the more transient nature of delight.

Lastly, SR-based nervousness showed a half-life of 59.4 seconds (CI [55.9 63.0]), making it the second least persistent state after confusion, but its half-life remained comparable to that of the self-reported engaged concentration and frustration. Nervousness was not measured either with our BROMP-based detectors, nor in either of the previous studies, so further comparisons are not possible.

Overall, we see a pattern in the current data where BROMP-based detectors have longer half-life values for boredom and engaged concentration, and much shorter values for other detectors. In contrast, our SR-based detectors have higher values for boredom than other detectors—though not to the extreme seen in either our BROMP-based detectors or those reported in Botelho et al. Instead, the values for our SR-based detectors sit between the BROMP-based detectors in both studies and the values reported for both types of detectors in D'Mello & Graesser, demonstrating considerable variability in half-life values, a finding that deserves further consideration.

4.2 Exponential Decay: Effects of Self-Efficacy & Situational Interest

Finally, we examined how different levels of self-efficacy (SE) and situational interest (SI) influence the decay rate of affective states in this dataset. We first report on the episodic differences in this data and then upon the half-life results.

4.2.1 Episodic Analysis

Table 2 presents the average number of affective episodes (i.e., the number of times an affective state appeared in a consecutive series of clips) per student, based on both their self-efficacy and interest levels and on the detectors making those predictions. In line with the shorter half-life values seen among the SR-based detectors in the previous section (Table 1), the SR-based detectors tended to identify more episodes of each affective state than the BROMP-based detectors (Mean difference of 1.8 episodes, CI [1.3, 2.4]). In other words, when affect was detected using SR-based detectors, it was more volatile—showing transitions from one affective state to another on a more frequent basis.

In addition to the differences in the number of episodes between the two suites of detectors, there were also differences in the number of episodes seen among the High vs. Low SE and High vs. Low SI groups. Notably, BROMP-based detectors found no episodes of boredom for both High SI and High SE students and no episodes of delight for the Low SE group.

Table 2. Average episodes of each affective state per student.

Method	Group	N	Bor	Conf	Eng	Fru	Del	Nerv
SR	High SE	18	3.7	3.3	5.6	1.7	2.9	0.3
	Low SE	18	3.6	2.3	4.8	1.1	2.6	0.2
	High SI	20	3.1	3.2	5.8	1.2	3.3	0.2
	Low SI	16	4.3	2.8	4.7	1.2	1.8	0.8
BROMP	High SE	18	0	1.5	3.3	1.1	0.4	-
	Low SE	18	0.3	1.7	2.9	0.6	0	-
	High SI	20	0	1.9	4.1	0.9	0.6	-
	Low SI	16	0.3	1.8	3.8	1.0	0.3	-

4.2.2 Exponential Decay—Motivational Differences

We next analyze how students' motivational measures impact the half-life values predicted by these detectors. Results for the SR-based detectors are presented in Table 3 and Figures 6 and 8, while results for the BROMP-based detectors are shown in Table 4 and Figure 7. As with our analysis above (where we compared the full data set to previous research), we present these results by each affective state. The goal is to better understand the relationships between these motivational constructs and the duration of each affective state, as that is now theorized as an important component of understanding affect dynamics.

Table 3. SR-based detector half-life estimates (seconds) for high/low levels of Situational Interest (SI) and Self-Efficacy (SE).

		Low SE/SI Group		High SE/SI Group		
		HL	**Range**	**HL**	**Range**	**hi-lo**
Boredom	SE	78.9	(69.9-88.5)	75.7	(66.7-85.2)	-3.2
	SI	97.6	(85.6-110.4)	99.2	(87.5-111.7)	1.6
Confusion	SE	50.2	(44.1-56.5)	41.8	(36.0-47.8)	-8.4
	SI	62.8	(55.8-70.0)	49.7	(43.3-56.5)	-13.0
Eng Conc	**SE**	**45.4**	**(38.4-52.7)**	**74.8**	**(65.8-84.2)**	**29.4**
	SI	**45.3**	**(37.6-53.4)**	**71.7**	**(61.5-82.3)**	**26.4**
Frustration	**SE**	**39.1**	**(32.2-46.5)**	**63.6**	**(53.0-74.8)**	**24.5**
	SI	80.4	(68.1-93.6)	92.2	(77.6-107.7)	11.8
Happiness	**SE**	**48.0**	**(41.8-54.6)**	**62.8**	**(54.9-70.7)**	**14.7**
	SI	**50.4**	**(45.5-55.5)**	**69.7**	**(64.0-75.6)**	**19.3**
Nervousness	**SE**	**38.9**	**(32.9-45.1)**	**139.3**	**(124.4-155.0)**	**100.4**
	SI	**37.6**	**(33.2-42.1)**	**11.0**	**(8.5-13.6)**	**-26.6**

***Grayscale indicates non-overlapping 95% credible intervals comparing high vs low SE/SI groups.**

In our first analysis, boredom showed the longest half-life values among all detectors, though that finding was much stronger for BROMP-based detectors than for SR-based detectors. As Table 4 and Figure 7 show, we can see that the BROMP-based findings were driven exclusively by students with low self-efficacy and/or low situational interest, who experienced episodes of boredom lasting longer than nine minutes. When measured with BROMP-based detectors, neither the high self-efficacy or high situational interest groups exhibited any boredom at all. However, this effect is not seen in the SR-based detectors (Table 3 and Figure 6), where there was minimal difference in the duration of boredom between both situational interest groups (97.6 and 99.2 seconds) and between both self-efficacy groups (75.7 and 78.9 seconds). These half-lives—around a minute and a half—differ substantially from the nearly nine minutes observed for the BROMP-based detector, a trend seen across all affective states.

Table 4. BROMP-based detector half-life estimates (seconds) across low/high levels of Situational Interest (SI) and Self-Efficacy (SE).

		Low SE/SI Group		High SE/SI Group		
		HL	**Range**	**HL**	**Range**	**hi-lo**
Boredom	**SE**	**602.8**	**(505.4-724.3)**	**-**	**(---)**	**-600.3**
	SI	**536.6**	**(435.6-668.0)**	**-**	**(---)**	**-533.1**
Confusion	**SE**	**57.4**	**(52.6-62.4)**	**38.1**	**(34.4-41.9)**	**-19.3**
	SI	**62.5**	**(58.1-67.0)**	**37.3**	**(34.0-40.6)**	**-25.2**
Eng Conc	**SE**	**216.4**	**(198.4-235.9)**	**302.1**	**(274.1-333.2)**	**85.7**
	SI	**183.6**	**(168.5-199.8)**	**282.2**	**(256.2-310.7)**	**98.6**
Frustration	**SE**	**58.7**	**(53.0-64.6)**	**86.5**	**(79.1-94.3)**	**27.8**
	SI	82.4	(75.6-89.5)	79.9	(73.4-86.6)	-2.5
Delight	**SE**	**-**	**(---)**	**22.4**	**(18.0-27.1)**	**22.6**
	SI	**35.3**	**(30.0-40.7)**	**23.6**	**(19.7-27.7)**	**-11.7**

***Grayscale indicates non-overlapping 95% credible intervals comparing high vs low SE/SI groups.**

We next examine the relationship between motivational measures and the half-life values of engaged concentration, which had overall half-life values of four and half minutes in the BROMP-based data but less than a minute in the SR-based data. In this analysis, we see that there are significant differences in both motivational measures for both BROMP and SR-based detectors. In the BROMP data, both high self-efficacy and situational interest groups tend to persist in being concentrated for periods around 5 minutes (302.1 and 282.2 sec.), which is around a minute and a half longer than students with low self-efficacy (216.4 sec.) and low situational interest (183.6 sec.). In the SR-based data, these differences are smaller, but in the same direction, students with high situational interest and high self-efficacy tended to remain in the engaged concentration state for nearly 30 seconds longer than their low self-efficacy and low situational interest peers (SE: 74.8 vs. 45.4 sec and SI: 71.7 vs. 45.3 sec.), suggesting a difference over 25 seconds between both groups (non-overlapping credible intervals). A similar result is observed according to the BROMP-based detectors.

In our first analysis, confusion had the shortest half-life values for both suites of detectors (under one minute), at only 39.9 seconds in the BROMP data and 50.6 seconds in the SR data. In the BROMP-based data (Table 4), this low half-life appears to be driven by students with high self-efficacy and high situational interest, whose values approached the half-minute mark (38.1 sec. CI [34.4, 41.9] and 37.3 sec., CI [34.0, 40.6]). Students with lower motivational measures tended to remain in the confusion state for nearly a minute or more (SE: 57.4 sec., CI [52.6, 62.4] and SI: 62.6 sec., CI [58.1, 67.0]). These differences are lower than those seen in the BROMP-based detectors for motivational differences in engaged concentration (approx. 20-25 sec. for confusion vs. 85-100 sec. for engaged concentration), but account for a greater proportion of the variance related to the estimated half-life across all students (39.9 sec. for confusion and 259.7 sec. for engaged concentration). A similar pattern was observed using the SR-based detectors, though, as Table 3 shows, the credible intervals overlapped in this comparison.

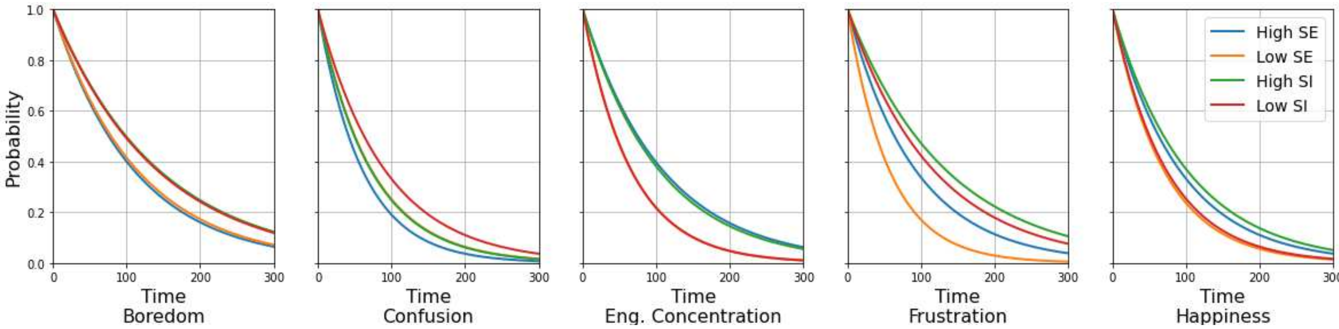

Figure 6. Half-life of SR-based detectors by high and low self-efficacy (SE) and situational interest (SI).

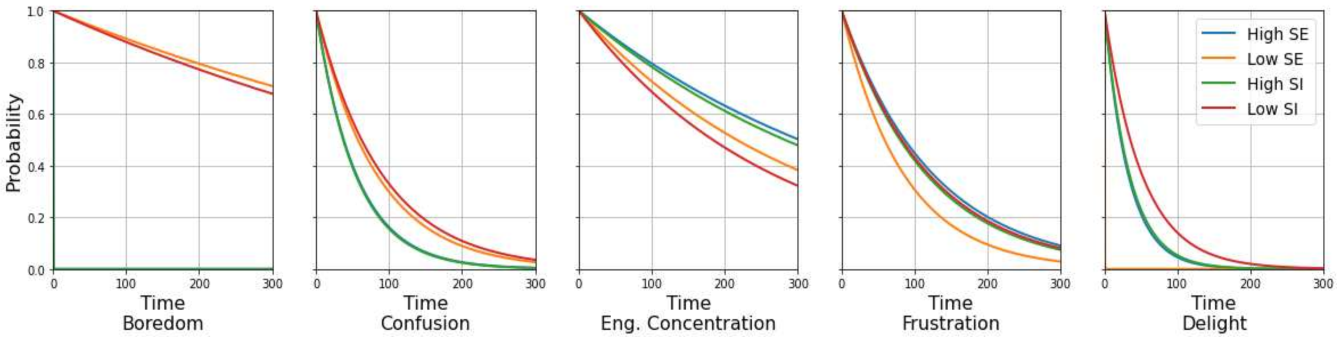

Figure 7. Half-life of BROMP-based detectors affective states by high and low self-efficacy (SE) and situational interest (SI).

Frustration's half-life values had highly consistent patterns within our own results and also in Botelho's results. Interestingly, in this analysis, neither suite of detectors showed differences based on situational interest, both of which remained within 79.9 to 92.2 seconds. All the estimated half-lives for the High and Low SI groups were greater than the overall results (SR: 65.7 sec.; BROMP: 74.3 sec), indicating that mid-level situational interest students tended to show shorter episodes of frustration according to both suites of detectors. Self-efficacy, however, had more contrasting results between the high and low groups. For both suites of detectors, students with high self-efficacy showed longer episodes of frustration (63.6 and 86.5 seconds for SR and BROMP-based detectors, respectively), persisting 20 seconds longer than the low self-efficacy groups (39.1 and 58.7 seconds for SR and BROMP).

Happiness and delight tended to rank in the middle compared to the estimated half-lives of the other affective states, with happiness lasting longer than delight. Students in the high self-efficacy and high situational interest groups also tended to remain in a state of happiness for longer durations (62.8 and 69.7 sec., respectively), according to the SR-based detectors. However, this difference (14.7 and 19.3 sec., respectively) was smaller than the differences observed for engaged concentration and frustration, particularly for self-efficacy. Notably, students with low self-efficacy either never experienced delight or, if they did, transitioned almost immediately to another affective state, similar to the boredom dynamics observed in the high situational interest and high self-efficacy groups. For situational interest, the pattern was the opposite. Students with high situational interest remained in the delight state for only 23.6 seconds, compared to 35.3 seconds for their low situational interest peers. However, this result merely indicates that high situational interest students transitioned out of delight more quickly. Since they also experienced twice as many episodes of delight as their low situational interest peers, this should not be interpreted as a negative association between situational interest and delight.

Finally, we examine the results for nervousness (Figure 8), which was the affective state with the second fastest decay for the SR-based detectors. Notably, the difference in nervousness between SE groups are greater than the difference for SI groups. Surprisingly, however, high self-efficacy is associated with higher half-life values for nervousness (139.3 sec., CI [124.4, 155.0]). In contrast, students with high situational interest have substantially shorter half-lives compared to those with low situational interest (11.0 sec. CI [8.5, 13.6]; CI [33.2, 42.1], respectively), though both situational interest groups have half-life values for nervousness that are well below the estimate we gave in the previous section for the full group (59.4 sec., CI [55.9, 63.0]).

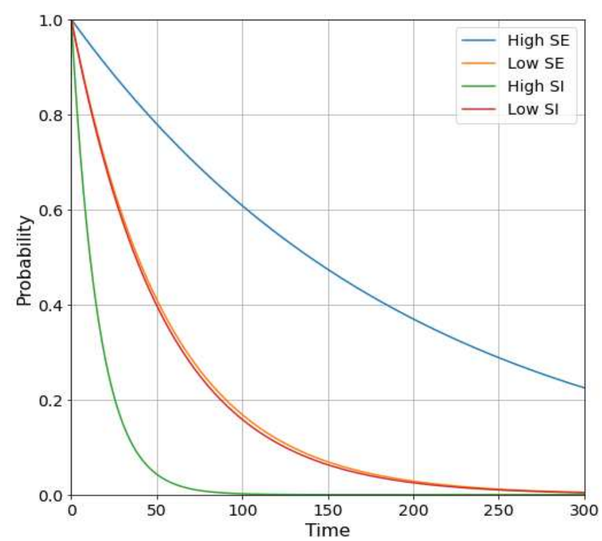

Figure 8. Half-life of nervousness among students with varying levels of self-efficacy and situational interest.

Overall, the BROMP-based detectors show greater sensitivity to motivational differences than the SR-based detectors. Although neither type of detector revealed differences in the relationship between situational interest and frustration, the BROMP-based detectors showed distinctions for all other motivational and affective combinations. In contrast, the SR-based detectors did not find self-efficacy or situational interest differences for either boredom or confusion. This suggests that if affect detectors are to be used in motivational interventions, BROMP-based detectors may offer some advantages, though notably, the SR-based detectors do offer the ability to detect nervousness, which also showed motivational differences.

5. DISCUSSION

In this study, we examined the half-lives of epistemic emotions using two different suites of detectors—one trained on in-game self-reported data and the other on BROMP observations. We also analyzed how the decay rates of these affective states interact with motivational factors such as self-efficacy and situational interest.

5.1 Comparison to Previous Results

Notably, the order and broader distribution of half-lives observed in our BROMP-based detectors (ranging 40 to 320 seconds) closely aligned with the findings of Botelho et al. [7], who also used BROMP-based detectors. In contrast, the SR-based detectors produced shorter half-lives, ranging between 50 and 100 seconds. In other words, the SR-based detectors portray students' affect dynamics as more volatile than the BROMP-based detectors, which show fewer transitions from one affective state to the next. Despite these differences, across both suites of detectors, boredom consistently exhibited the slowest decay rate (highest persistence), while confusion had the fastest decay rate (lowest persistence).

Additionally, the half-lives observed in our study were substantially higher than those reported by D'Mello and Graesser [13]. It is not clear why D'Mello & Graesser's students appeared to experience more volatility in their affective states, but it could be related to methodological differences, as their transitions were manually labeled on intervals of 20 seconds by human experts and students, and the data was collected within a laboratory study, which may have impacted how they approached the task.

The strong similarity between our results and those reported by Botelho et al., despite their use of a very different non-game learning platform, suggests some degree of generalizability in the typical half-lives of affective states. However, the sharp contrast with the findings from D'Mello & Graesser also indicates that differences in estimated half-lives may not only reflect variations in learning contexts but may also stem from similarities or differences in the methodologies used across these studies. Specifically, four distinct methods were employed to determine students' affective states: (1) self-reports from students (this study), (2) retrospective self-judgments after watching recordings of themselves (D'Mello & Graesser), (3) real-time observations by human evaluators (this study and Botelho), and (4) expert labeling based on video reviews of students (D'Mello & Graesser).

Each of these methods for obtaining ground truth present unique challenges and may capture slightly different signals. The implications of these differences could help to explain the patterns we are reporting upon here. For example, while the student is the only person with direct access to their own emotions, the way they categorize affective states can vary depending on whether they report them in real time or recall and label them later while watching a video of their past experiences. When reviewing a video, students may interpret their facial expressions as signs of emotional transitions, potentially perceiving shifts in affect that felt more gradual in the moment. For instance, a person who feels bored or has low interest in a particular activity might be more likely to report boredom in real time than they would be if, after that experience is resolved positively, they are asked to reflect on it later. If their overall perception of the educational experience is positive at the end, they may rely on that perception to identify fewer instances of boredom and instead report more instances of delight or engaged concentration than they would have labeled in real-time. In other words, the opportunity to self-reflect might influence their reporting in ways that do not capture the *in situ* experience.

Similar differences may arise between real-time, in-person observations and retrospective assessments by trained experts. The additional contextual information that might help determine a student's affective state more accurately, cannot be fully captured in video. On the other hand, since it is impractical to observe all students at each instant—just as students should not be asked to report their affective state every 20 seconds—video recordings might capture more granular and subtle expressions associated with more momentary states. In contrast, a real-time observer who is not permanently assessing the same student (as doing so would disrupt the learning experience) may perceive a slightly more stable emotional signal rather than capturing every fleeting shift in affect.

In addition to differences in the collection of the ground truth, another important distinction between D'Mello & Graesser, and Botelho et al.'s and our study lies in the use of detectors. Because detectors estimate general behavioral patterns, they may capture more stable emotional signals over time. In contrast, labeling based on video recordings—without the use of detectors—may focus more on fleeting affective states, such as brief facial expressions lasting only a few seconds (e.g., 0–5 seconds), which may not fully represent the predominant emotional state over longer time intervals (e.g., 20–30 seconds).

Finally, the choice of a 1-minute window (D'Mello & Graesser) versus a 5-minute window (Botelho et al. and our study) can significantly impact the estimation of decay rates and half-lives. While the exponential function provides a good approximation of how the probability of remaining in a given affective state declines over time, as seen in D'Mello & Graesser's and Botelho et al., studies [7, 13], it is not a perfect model. Shorter windows (e.g., 1 minute) may be particularly useful for capturing the initial decay of an emotion or tracking affective states that are brief (e.g., delight). In contrast, longer windows (e.g., 5 minutes) may be better suited for capturing both the initial decay and the later-stage decline of more persistent states (e.g., boredom or engaged concentration).

These methodological differences and the potential limitations of each approach do not imply that any of these studies are incorrect. Rather, they highlight how research design choices can shape results. The most appropriate method depends on the type of affective signal researchers aim to capture—whether momentary emotional states, students' real-time self-perceptions, more stable affective states, or external observers' interpretations—all of which might be correlated with important learning outcomes (e.g., [37]). Future research should focus on identifying the specific aspects of the affective constellation that each labeling method captures to develop a more nuanced understanding of the implications of half-lives estimated across multiple ground truths and contexts.

5.2 Alignment to SDVET Model

Our analysis of the differences across groups with different levels of self-efficacy and situational interest allows us to explore some of the theoretical and empirical claims made by the SDVET model [25]. Although we did not have a large enough sample to run an analysis on prior knowledge or time in game, which are important components of the SDVET predictions, situational interest is a reasonable proxy for value and self-efficacy is explicitly included in the model.

In particular, we see that engaged concentration episodes have significantly longer half-life values for students with high situational interest, corresponding to the SDVET model. This is true regardless of detector type. Although there were no SDVET predictions for the effect of self-efficacy on engaged concentration, the effects of high self-efficacy were the same as those seen with situational interest.

Similarly, the longer half-life of confusion among students with low situational interest supports the SDVET model's prediction that this affective state is most likely to occur when value is low. However, this finding appears only in the BROMP-based data, where low self-efficacy follows the same pattern.

The BROMP-based data also align with SDVET model predictions for boredom. Specifically, no episodes of boredom were observed among students with high situational interest or high self-efficacy. This supports the SDVET prediction that higher self-efficacy enables students to persist longer before cycling between frustration and boredom.

Because our data collection did not distinguish between canonical frustration and pleasurable frustration [17], our results around frustration are more difficult to interpret. Students with low self-efficacy are predicted in SDVET to transition more quickly into the space where cycles between canonical frustration and boredom occur. Students with high situational interest are predicted by SDVET to spend more time in pleasurable frustration, and higher self-efficacy would be predicted to further extend those experiences. In our results, students with high self-efficacy are more likely to experience longer episodes of frustration, which would be predicted if these experiences were pleasurable. To the degree that there is considerable diversity in how confusion and frustration manifest [2, 9], it may be worth considering how to capture different forms of confusion and frustration in future studies of this nature. One approach could involve follow-up questions administered shortly after students self-report or are observed experiencing confusion or frustration. These prompts could ask about the cause of the emotion, whether the issue was resolved, and how the student is feeling now, providing deeper insight into the underlying affective dynamics.

Although there were no predictions for delight in the SDVET model, the relationship between self-efficacy and delight in this study would be compatible with the idea that high self-efficacy is associated with pleasurable frustration. Notably, no student in the low self-efficacy group showed any episodes of delight (in the BROMP-based data), and although happiness (the SR-based equivalent) does occur among low self-efficacy students, the duration of this emotion is longer among those with high self-efficacy.

That said, this interpretation is complicated by results related to situational interest, which contrasts with the relationship between self-efficacy and delight/happiness. For the SR-based happiness detector, low situational interest is associated with shorter bouts of happiness, but for the BROMP-based delight detector, the results are the opposite; students with low situational interest experienced longer bouts of delight. However, despite persisting longer in this state, students with low situational interest experienced only half as many episodes of delight as their high situational interest peers. This suggests that, overall, they do not necessarily experience more delight.

One possible explanation is that students who are already interested in a specific domain may become less sensitive to novelty, surprise, or success after overcoming challenges—factors that typically trigger and prolong the high-intensity emotion of delight [27]. As a result, their experiences of delight may occur more frequently but be shorter in duration. In contrast, students with high situational interest tend to persist longer in other positive but less intense states [24, 31], such as happiness and engaged concentration, as observed in this study. Overall, these students appear to have a more positive—but potentially less intense—experience of the game. Further research exploring the underlying causes of delight, happiness, and engaged concentration may help validate this interpretation.

5.3 Potential for Interventions

The strong similarity between Botelho's findings and ours, even using different systems (a question-based learning platform and an educational game), suggests the presence of general trends in affective half-life that could inform interventions across multiple platforms. The high persistence of boredom (lasting over five minutes according to the BROMP-based detectors) indicates that this affective state may be particularly difficult to overcome, a finding previously noted [23]. For this reason, researchers should focus on predicting boredom before it occurs, as students may struggle to transition out of it once they are bored [36]. This prolonged persistence in this affective state can negatively impact multiple learning outcomes [5].

Additionally, understanding the half-lives of confusion and frustration—affective states that are not inherently negative but can lead to undesirable outcomes if unresolved [15, 28, 29, 34]—can help determine the optimal time frame for interventions. For example, educational systems or games could use this information to provide timely hints that assist students in overcoming these states. Interventions should not necessarily be immediate, as confusion and frustration can contribute to positive outcomes [14, 16, 20, 22, 26]. However, waiting until the students potentially transition to boredom might also have negative effects. Therefore, knowing that there is a window of 30 seconds to a minute in which these affective states can be effectively resolved—potentially allowing students to transition back to engaged concentration [22]—enables systems to deliver more strategically timed interventions. Still, as noted earlier, different forms of confusion and frustration may vary in duration, impact on learning, and the types of interventions they require [2, 9].

The interplay between motivational factors and the half-lives of affective states also plays a crucial role in determining the most effective timing for interventions. For example, as proposed by the SDVET framework, students with higher self-efficacy or situational interest may be better equipped to manage frustration, making them more likely to self-regulate and transition back to engaged concentration even if they persist in frustration for longer [25]. This hypothesis is further supported by the absence of boredom detected in these students. In contrast, students with lower situational interest or self-efficacy may persist in frustration for a shorter duration but, in some cases, transition to boredom more quickly, which could lead to the potential negative consequences associated with this affective state. Further research is necessary to evaluate this hypothesis.

The association between confusion and motivational factors appears to be the opposite of what was observed for frustration. Students with low situational interest and self-efficacy tend to persist in confusion for longer, suggesting that they may require more time to resolve their confusion and either return to engaged concentration or transition to another affective state. Based on this, further analysis is needed to determine when these students are able to self-regulate, overcome confusion, and re-engage independently versus when external support from the system or game is necessary to facilitate their learning process.

Additionally, the shorter half-life of engaged concentration—generally the most common emotion for learners using digital learning platforms [19]—in students with low situational interest and self-efficacy suggests that their affective dynamics may be more volatile, causing them to transition out of the flow state more quickly and frequently. These findings align with predictions in the SDVET framework [24] and highlight the importance of predicting affective states in this group to develop targeted interventions, such as motivational messages [11], that could help them sustain engagement for longer periods.

6. CONCLUSIONS

This study highlights the importance of understanding the half-lives of affective states as a crucial step toward better understanding students' affective dynamics and designing more effective interventions in learning environments. Our findings suggest that while emotions like boredom and engaged concentration tend to persist for extended periods, emotions like confusion have shorter durations. The alignment between our results and previous research investigating a different digital non-game learning environment indicates the presence of general affective trends that could inform adaptive learning systems across different platforms. However, there is a need to replicate these results in other learning contexts and different domains. Additionally, our findings show that motivational factors, such as self-efficacy and situational interest, influence student persistence in these states. As this is, to our knowledge, the first study of its kind, further replication across varied educational settings—incorporating a range of motivational measures—is warranted.

Overall, the results presented in this study could guide the development of timely interventions aimed at preventing boredom while promoting positive transitions between confusion, frustration, happiness, and engaged concentration—pathways that may lead to improved learning outcomes. These insights underscore the importance of considering both motivational factors and the appropriate time frame when designing affect-based interventions.

7. 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.

8. REFERENCES

[1] Andres, J.Ma.A.L., Ocumpaugh, J., Baker, R.S., Slater, S., Paquette, L., Jiang, Y., Karumbaiah, S., Bosch, N., Munshi, A., Moore, A. and Biswas, G. 2019. Affect Sequences and Learning in Betty's Brain. *Proceedings of the 9th International Conference on Learning Analytics & Knowledge* (New York, NY, USA, 2019), 383–390.

[2] Baker, R.S., Cloude, E., Andres, J.M.A.L. and Wei, Z. 2025. The Confrustion Constellation: A New Way of Looking at Confusion and Frustration. *Cognitive Science*. 49, 1 (Jan. 2025), e70035. DOI:https://doi.org/10.1111/cogs.70035.

[3] Baker, R.S., Gowda, S.M., Wixon, M., Kalka, J., Wagner, A.Z., Salvi, A., Aleven, V., Kusbit, G.W., Ocumpaugh, J. and Rossi, L. 2012. Towards Sensor-Free Affect Detection in Cognitive Tutor Algebra. *International Conference on Educational Data Mining* (2012), 126–133.

[4] Baker, R.S., Ocumpaugh, J.L. and Andres, J. 2020. BROMP quantitative field observations: A review. *Learning Science: Theory, Research, and Practice. New York, NY: McGraw-Hill*. (2020), 127–156.

[5] Baker, R.S.J. d., D'Mello, S.K., Rodrigo, Ma.M.T. and Graesser, A.C. 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 (Apr. 2010), 223–241. DOI:https://doi.org/10.1016/j.ijhcs.2009.12.003.

[6] Baker, R.Sj. and Ocumpaugh, J. 2014. Interaction-Based Affect Detection in Educational Software. *The Oxford handbook of affective computing*. (2014), 233.

[7] Botelho, A.F., Baker, R.S., Ocumpaugh, J. and Heffernan, N.T. 2018. Studying Affect Dynamics and Chronometry Using Sensor-Free Detectors. *International Conference on Educational Data Mining* (2018), 157–166.

[8] Britner, S.L. and Pajares, F. 2006. Sources of science self-efficacy beliefs of middle school students. *Journal of Research in Science Teaching*. 43, 5 (2006), 485–499. DOI:https://doi.org/10.1002/tea.20131.

[9] Cloude, E.B., Munshi, A., Andres, J.M.A., Ocumpaugh, J., Baker, R.S. and Biswas, G. 2024. Exploring Confusion and Frustration as Non-linear Dynamical Systems. *Proceedings of the 14th Learning Analytics and Knowledge Conference* (New York, NY, USA, 2024), 241–252.

[10] Csikszentmihalyi, M. 2000. *Beyond boredom and anxiety*. Jossey-bass.

[11] DeFalco, J.A., Rowe, J.P., Paquette, L., Georgoulas-Sherry, V., Brawner, K., Mott, B.W., Baker, R.S. and Lester, J.C. 2018. Detecting and Addressing Frustration in a Serious Game for Military Training. *International Journal of Artificial Intelligence in Education*. 28, 2 (Jun. 2018), 152–193. DOI:https://doi.org/10.1007/s40593-017-0152-1.

[12] D'Mello, S. and Graesser, A. 2012. Dynamics of affective states during complex learning. *Learning and Instruction*. 22, 2 (2012), 145–157. DOI:https://doi.org/10.1016/j.learninstruc.2011.10.001.

[13] D'Mello, S. and Graesser, A. 2011. The half-life of cognitive-affective states during complex learning. *Cognition and Emotion*. 25, 7 (Nov. 2011), 1299–1308. DOI:https://doi.org/10.1080/02699931.2011.613668.

[14] D'Mello, S., Lehman, B., Pekrun, R. and Graesser, A. 2014. Confusion can be beneficial for learning. *Learning and Instruction*. 29, (2014), 153–170. DOI:https://doi.org/10.1016/j.learninstruc.2012.05.003.

[15] D'Mello, S.K., Lehman, B. and Person, N. 2010. Monitoring Affect States During Effortful Problem Solving Activities. *International Journal of Artificial Intelligence in Education*. 20, 4 (2010), 361–389. DOI:https://doi.org/10.3233/JAI-2010-012.

[16] Fancsali, S. 2014. Causal discovery with models: behavior, affect, and learning in cognitive tutor algebra. *Educational Data Mining 2014* (2014).

[17] Gee, J.P. 2012. *Situated language and learning: A critique of traditional schooling*. Routledge.

[18] Heffernan, N.T. and Heffernan, C.L. 2014. The ASSISTments Ecosystem: Building a Platform that Brings Scientists and Teachers Together for Minimally Invasive Research on Human Learning and Teaching. *International Journal of Artificial Intelligence in Education*. 24, 4 (Dec. 2014), 470–497. DOI:https://doi.org/10.1007/s40593-014-0024-x.

[19] Karumbaiah, S., Baker, R.S., Ocumpaugh, J. and Andres, J.Ma.A.L. 2023. A Re-Analysis and Synthesis of Data on Affect Dynamics in Learning. *IEEE Transactions on Affective Computing*. 14, 2 (2023), 1696–1710. DOI:https://doi.org/10.1109/TAFFC.2021.3086118.

[20] Lehman, B., D'Mello, S., Strain, A., Mills, C., Gross, M., Dobbins, A., Wallace, P., Millis, K. and Graesser, A. 2013. Inducing and Tracking Confusion with Contradictions during Complex Learning. *International Journal of Artificial Intelligence in Education*. 22, 1–2 (2013), 85–105. DOI:https://doi.org/10.3233/JAI-130025.

[21] Linnenbrink-Garcia, L., Durik, A.M., Conley, A.M., Barron, K.E., Tauer, J.M., Karabenick, S.A. and Harackiewicz, J.M. 2010. Measuring situational interest in academic domains. *Educational and psychological measurement*. 70, 4 (2010), 647–671.

[22] Liu, Z., Pataranutaporn, V., Ocumpaugh, J. and Baker, R. 2013. Sequences of frustration and confusion, and learning. *International Conference on Educational Data Mining* (2013), 114–120.

[23] Nasiar, N., Zambrano, A.F., Ocumpaugh, J., Goslen, A., Rowe, J., Vandenberg, J., Esiason, J. and Hutt, S. 2024. The Influence of Different Measurement Approaches on Student Affect Transitions Using Ordered Networks. *Advances in Quantitative Ethnography* (Cham, 2024), 195–203.

[24] Ocumpaugh, J., Baker, R.S. and Rodrigo, M.M.T. 2015. Baker Rodrigo Ocumpaugh Monitoring Protocol (BROMP) 2.0 Technical and Training Manual. (2015).

[25] Ocumpaugh, J., Nasiar, N., Zambrano, A.F., Goslen, A., Vandenberg, J., Esiason, J., Rowe, J. and Hutt, S. 2025. Refocusing the lens through which we view affect dynamics: The Skills, Difficulty, Value, Efficacy and Time Model. (Dublin, Ireland, 2025).

[26] Pardos, Z.A., Baker, R.S., San Pedro, M.O., Gowda, S.M. and Gowda, S.M. 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.

[27] Pekrun, R. 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:https://doi.org/10.1007/s10648-006-9029-9.

[28] Richey, J.E., Andres-Bray, J.M.L., Mogessie, M., Scruggs, R., Andres, J.M.A.L., Star, J.R., Baker, R.S. and McLaren, B.M. 2019. More confusion and frustration, better learning: The impact of erroneous examples. *Computers & Education*. 139, (2019), 173–190. DOI:https://doi.org/10.1016/j.compedu.2019.05.012.

[29] Richey, J.E., McLaren, B.M., Andres-Bray, M., Mogessie, M., Scruggs, R., Baker, R. and Star, J. 2019. Confrustion in Learning from Erroneous Examples: Does Type of Prompted Self-explanation Make a Difference? *Artificial Intelligence in Education* (Cham, 2019), 445–457.

[30] Rowe, J.P., Shores, L.R., Mott, B.W. and Lester, J.C. 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:https://doi.org/10.3233/JAI-2011-019.

[31] Russell, J.A. 1980. A circumplex model of affect. *Journal of personality and social psychology*. 39, 6 (1980), 1161.

[32] Sabourin, J.L. and Lester, J.C. 2013. Affect and engagement in Game-BasedLearning environments. *IEEE transactions on affective computing*. 5, 1 (2013), 45–56.

[33] Sawyer, R. and Rowe, J. Filtered Time Series Analyses of Student Problem-Solving Behaviors in Game-based Learning.

[34] Schneider, B., Krajcik, J., Lavonen, J., Salmela-Aro, K., Broda, M., Spicer, J., Bruner, J., Moeller, J., Linnansaari, J., Juuti, K. and Viljaranta, J. 2016. Investigating optimal learning moments in U.S. and finnish science classes. *Journal of Research in Science Teaching*. 53, 3 (Mar. 2016), 400–421. DOI:https://doi.org/10.1002/tea.21306.

[35] Wixon, M., Arroyo, I., Muldner, K., Burleson, W., Rai, D. and Woolf, B. 2014. The opportunities and limitations of scaling up sensor-free affect detection. *International Conference on Educational Data Mining* (2014), 145–152.

[36] Zambrano, A.F., Baker, R.S., Baral, S., Heffernan, N.T. and Lan, A. 2024. From Reaction to Anticipation: Predicting Future Affect. *Proceedings of the 17th International Conference on Educational Data Mining* (Atlanta, Georgia, USA, Jul. 2024), 566–574.

[37] Zambrano, A.F., Nasiar, N., Ocumpaugh, J., Goslen, A., Zhang, J., Rowe, J., Esiason, J., Vandenberg, J. and Hutt, S. 2024. Says Who? How different ground truth measures of emotion impact student affective modeling. *Proceedings of the 17th International Conference on Educational Data Mining* (Atlanta, Georgia, USA, Jul. 2024), 211–223.